(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number
WO 02/02599 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/EP01/07532

(22) International Filing Date: 2 July 2001 (02.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/215,759 30 June 2000 (30.06.2000) US

(71) Applicant *(for all designated States except US)*: **INGENIUM PHARMACEUTICALS AG** [DE/DE]; Fraunhofer Strasse 13, 82152 Martinsried (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **Wattler, Frank** [DE/DE]; Bennostrasse 11A, 82131 Stockdorf (DE). **Wattler, Sigrid** [DE/DE]; Bennostrasse 11A, 82131 Stockdorf (DE). **Trommler, Paul** [DE/DE]; Randeckstrasse 4, 81375 München (DE). **Nehls, Michael, C.** [DE/DE]; Paul-Keller-Strasse 6, 82131 Stockdorf (DE).

(74) Agent: **DÖRRIES, FRANK-MOLNIA & POHLMAN**; Postfach 221661, 80506 Munich (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 02/02599 A2

(54) Title: HUMAN G PROTEIN-COUPLED RECEPTOR IGPcR17, AND USES THEROF

(57) Abstract: A novel human G protein-coupled receptor (GPcR) protein, IGPcR17, is identified and characterized. IGPcR17-encoding nucleotides, IGPcR17 proteins and fusion proteins, antibodies to the receptor, host cell expression systems, animal models in which the IGPcR17 gene is mutated, recombinant knock-out animals that do not express IGPcR17 and transgenic animals that express an IGPcR17 transgene are encompassed by the invention, as are compounds that modulate gene expression or receptor activity of IGPcR17 and their use for drug screening, and diagnosis or treatment of diseases, particularly central nervous system disorders.

# Human G protein-coupled receptor IGPcR17, and uses thereof

## Field of the Invention

The present invention relates to the field of cellular and molecular biology, protein biochemistry, and pharmacology. The invention relates particularly to the identification of the polynucleotide sequence of a novel G protein-coupled receptor (GPcR) and the characterization of nucleic acids that encode this G protein-coupled receptor, which is referred to herein as IGPcR17. The invention further relates to animal orthologs of the human gene encoding IGPcR17, to expression of both human and animal proteins, to the function of the gene product and to uses for the receptor, and its ligands in drug screening and in diagnosing, preventing and treating disease, particularly dysfunctions associated with signal processing in the central nervous system (CNS). Animal models of such diseases and dysfunctions, in which the IGPcR17 gene is mutated, knocked-out or present in the form of a transgene, are also incorporated within the invention.

## Background of the Invention

It is well established that many medically significant biological processes are mediated by proteins that participate in signal transduction pathways involving G proteins and second messengers; *e.g.* cAMP, diacylglycerol and inositol phosphates (Lefkowitz, 1991, Nature, 351:353-354). Herein these proteins are referred to as proteins participating in pathways with G protein-coupled receptors, either as the receptors themselves, such as those for adrenergic agents and dopamine (Kobilka, BK, *et al.*, 1987, P.N.A.S., USA, 84:46-50; Kobilka BK *et al.*, 1987, Science, 238:650-656; Bunzow JR, *et al.*, 1988, Nature, 336:783-787), or as the G proteins to

which the receptors are coupled, or as effector proteins, *e.g.* adenylate cyclase, protein kinase A and protein kinase C (Simon MI, *et al.*, 1991, Science, 252:802-808).

Upon hormone binding to a GPcR the receptor interacts with the heterotrimeric G protein and induces the dissociation of GDP from the guanine nucleotide-binding site. At normal cellular concentrations of guanine nucleotides, GTP fills the site immediately. Binding of GTP to the alpha subunit of the G protein causes the dissociation of the G protein from the receptor and the dissociation of the G protein into alpha and beta-gamma subunits. The GTP-carrying form then binds to the generator of an intracellular second messenger: in one common form of signal transduction, activated adenylate cyclase. Hydrolysis of GTP to GDP, catalyzed by the intrinsic GTPase activity of the G protein alpha subunit, returns the G protein to its basal, inactive form. The GTPase activity of the alpha subunit determines the time period during which the G protein is active. The GDP-bound form of the alpha subunit (alpha.GDP) has high affinity for the beta-gamma subunit complex and subsequent re-association of G protein subunits alpha.GDP with beta-gamma returns the G protein to the basal state. Thus the G-protein serves a dual role: as an intermediate that relays the signal from receptor to effector (in this example adenylate cyclase), and as a timer that controls the duration of the signal.

Examples of members of the G protein-coupled receptor family gene family include acetylcholine, adenosine, adrenergic, bradykinin, cAMP, calcitonin, capsaicic, CCK, CGRP, CRF, cytomegalovirus, dopamine, endothelial differentiation gene-1, endothelin, FSH, galanin, histamine, kinin, motilin, muscarinic, neurokinin, neuropeptideY, neurotensin, nociceptin, odorant, opsin, rhodopsin, serotonin, somatostatin, thrombin, TSH and VIP receptors. Alteration of GPcR genes and gene products can cause medical disorders, dysfunctions, or diseases hereafter generally referred to as "diseases". The mechanism of disease may be due to a loss of receptor function or by constitutive receptor activation (reviewed by Coughlin *et al.*, 1994, Curr. Opin. Cell Biol., 6:191-197). For example, activating mutations of rhodopsin

receptor have been found in *retinitis pigmentosa* and congenital night blindness (Rao *et al.*, 1994, Nature 367:639-642); mutations of TSH receptor have been detected in sporadic and inherited hyperthyroidism (Parma *et al.*, 1993, Nature 365:649-651) and nephrogenic *diabetes insipidus* (Holtzman *et al.*, 1993, Hum. Mol. Genet. 2:1201-1204); mis-sense mutations in the luteinizing hormone receptor (LHR) gene, leading to constitutive activation of the LHR, have been shown to be associated with a condition in boys called familial male-limited precocious puberty (Cocco *et al.*, 1996, Hum. Mut., 7:164-166; Kosugi *et al.*, 1995, Hum. Mol. Genet., 4:183-188). Moreover, dopamine receptors are known to bind neuroleptic drugs used for treating CNS disorders.

As a characteristic feature, G protein-coupled receptors exhibit seven transmembrane domains which are connected by three hydrophilic extracellular loops alternating with three intracellular loops. Most G protein-coupled receptors have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed to stabilize functional protein structure. The seven transmembrane domains or regions are designated as TM1, TM2, TM3, TM4, TM5, TM6 and TM7. The cytoplasmic loop which connects TM5 and TM6 may be a mayor component of the G protein binding domain.

Most G protein-coupled receptors contain potential phosphorylation sites within the third cytoplasmic loop and/or the carboxyl terminus. For several GPcRs, such as the beta-adrenergic receptor, phosphorylation by protein kinase A and/or specific receptor kinases mediates receptor desensitization.

It has also been shown that certain G protein-coupled receptors, *e.g.* the calcitonin receptor-like receptor, might interact with small single pass membrane proteins called receptor-activity-modifying-proteins (RAMPs). This interaction of the GPcR with a certain RAMP determines which natural ligands have relevant affinity for the GPcR-RAMP combination and regulate the functional signaling activity of the complex (McLathie LM, *et al.*, 1998, Nature, 393:333-339).

For some receptors, the ligand binding sites of the G protein-coupled receptors are believed to comprise hydrophilic sockets formed by several GPcR transmembrane domains, said sockets being surrounded by hydrophobic residues of the G protein-coupled receptors. The hydrophilic side of each GPcR transmembrane helix is thought to face inward and form a polar ligand-binding site. TM3 has been implicated in several G protein-coupled receptors as having a ligand-binding site, such as the TM3 aspartate residues. TM5 serine residues, and TM6 asparagine and TM6 or TM7 phenylalanine or tyrosine residues are also implicated in ligand binding. G-protein coupled receptors bind to a variety of ligands ranging from small biogens to peptides, small proteins and large glycoproteins (Strader CD, *et al.*, 1994, Annu. Rev. Biochem., 63:101-132).

G protein-coupled receptors can be coupled intracellulularly by heterotrimeric G proteins to various intracellular enzymes, ion channels and transporters (see Johnson *et al.*, 1989, Endoc. Rev., 10:317-331). Different G protein alpha-subunits preferentially stimulate particular effectors to modulate various biological functions in a cell. Phosphorylation of cytoplasmic residues of G protein-coupled receptors has been identified as an important mechanism for the regulation of G protein coupling of some G protein-coupled receptors. G protein-coupled receptors are found in numerous sites within animal, and particularly mammalian hosts.

Evolutionary analyses suggest that the ancestor of G protein-coupled receptors originally developed in concert with complex body plans and nervous systems. With the exception of the visual opsins, the genes for the GPcR family have, in most instances, been characterized by a lack of introns within their coding sequences thus precluding the generation of receptor diversity through alternative splicing. Recent data support the idea that dimerization of G protein-coupled receptors is important in different aspects of receptor biogenesis and function. When considering, for example, the nervous system, the existence of homodimers and heterodimers of neurotransmitter G protein-coupled receptors offers an attractive explanation of the

great diversity and plasticity that is characteristic of such a highly organized and complex system (see Bouvier M, 2001, Nature Rev. Neuroscience, 2:274-286).

In order to understand the role of particular G protein-coupled receptors in normal physiology and disease, knock-out mice have been generated in which the endogenous genes encoding these receptors have been individually targeted. Studies of mas-protooncogene knock-out mice indicate that this GPcR is a determinant of heart rate and blood pressure variability (Walther *et al.*, 2000, Braz J, Med. Biol. Res., 33:1-9). Male mice showed increased anxiety, indicating a function for mas, which is an angiotensin receptor acting in the central nervous system (CNS) (Walther *et al.*, 1998, J. Biol. Chem., 273:11867-11873). Incerti *et al.*, (2000, Hum. Molec. Genet., 9:2871-2788) generated and characterized mice deficient in Oa1 (ocular albinism)-deficient mice by gene targeting. Ophthalmologic examination showed hypo-pigmentation of the ocular fundus in mutant animals compared with wildtype. Also demonstrated was a misrouting of the optic fibers at the chiasm and the presence of giant melanosomes in retinal pigment epithelium, as observed in OA1 patients. Prostaglandin E2 receptor knockout mice show a mild change in renal water handling, while EP2 receptor knock-out mice display salt-sensitive hypertension (Breyer *et al.*, 2000, Curr. Opin. Nephrol. Hypertens., 9:23-29).

Based on malfunctions discovered in signaling pathways several drugs have been developed, for example, a compound that blocks the farnesylation of ras as a tumor inhibitor, a JAK-2 blocker as an inhibitor of recurrent pre-B cell acute lymphoblastic leukemia, and a platelet-derived growth factor receptor kinase as a blocker of restenosis (Reviewed in Levitzki A, 1996, Curr. Opin. Cell Biol., 8:239-244). G protein-coupled receptors have been identified and successfully used as targets for several existing drugs; for example, dopamine and serotonin G protein-coupled receptors have been targeted for CNS diseases, angiotensin, muscarinic and adrenergic receptor G protein-coupled receptors have been targeted for cardiovascular diseases, histaminic G protein-coupled receptors have been targeted for respiratory diseases, the prostaglandin GPcR has been targeted for opthalmic

purposes, and calcitonin and estrogen for treatment of arthritis.

Recently, Hollopeter *et al.* described human IGPcR-17 as the P2Y12 platelet ADP receptor, encoding a prominent 1.0 kbp (kilo base pair) transcript in human platelet mRNA. The gene was isolated from a human platelet cDNA library and the protein is targeted by certain antithrombotic drugs, indicating that the receptor may be useful in the development of new antithrombotic drugs. Expression of this gene was also noted in numerous sub-regions of the brain, represented by a 2.4 kbp transcript (Hollopeter G, *et al.*, 2001, Nature, 409:202-207).

The following information relating to CNS function is provided in relation to the particular human G protein-coupled receptor, and its animal orthologs, that are disclosed by the present invention. The cerebral cortex of the brain is divided into four distinct sections, the lobes, which are associated with distinct functions: the frontal lobe includes the motor area and is also connected to thinking, planning and spoken language. The parietal lobe includes the somatosensory area. The temporal lobe is active in the processes involved in understanding language and it contains the auditory area. The occipital lobe includes the area responsible for vision. It is not known how the cerebral cortex processes or stores the innumerable sensory stimuli and perceptions that it receives and is able to mark them as memories that can subsequently be recalled from storage, but it is generally agreed that the large association areas of the cerebral cortex are the location of stored memories (reviewed by Duus P, 1998, "*Topical Diagnosis in Neurology*", 3rd edition, Thieme).

Astrocytes ("glial" cells) constitute nearly half of the cells in human brain. They ensheath synapses throughout the CNS and are thought of as synaptic support cells, clearing ions and neurotransmitters from the synaptic cleft (see for example Bergles DE, and Jahr CE, 1998, J. Neurosci., 18:7709). Recently Ullian *et al.* reported evidence from *in vitro* assays that exposure of CNS neurons to astrocytes is able to increase the number of mature, functional synapses on the CNS neurons by seven-fold in comparison to neurons remaining unexposed to astrocytes. In addition,

astrocytes are required for synaptic maintenance, since within a few days of removing astrocytes from neurons, the neurons begin to shed synapses (Ullian EM, *et al.*, 2001, Science, 291:657-661). *In vivo*, neurons send out dendrites and axons to the appropriate parts of the brain very early in development, but they do not form the majority of their synapses with other neurons until many days later – which is approximately the same time period at which astrocytes mature. Ullian *et al.* suggest that during normal embryonic development, initially formed synapses may be immature and highly plastic. The post-natal appearance of astrocytes increases the number of synapses and locks synaptic circuitry in place. According to Ullian *et al.*, glial cells may also play important roles in adult neural plasticity underlying learning and memory, including short-term memory and long-term memory: a hypothesis supported by many researchers convinced that learning forges stronger connections between neurons. The signal that astrocytes send to trigger synapse formation by neurons has not yet been identified.

Because of the vital role of G protein-coupled receptors in the communication between cells and their environment, such receptors are attractive targets for therapeutic intervention. G protein-coupled receptors have led to more than half of the currently known drugs (Drews, Nature Biotechnology, 1996, 14:1516). Mechanistically, approximately 50% to 60% of all clinically relevant drugs act by modulating the functions of various G protein-coupled receptors, as either agonist (activating activity) or antagonist (blocking activity) of a GPcR (Cudermann *et al.*, 1995, J. Mol. Med., 73:51-63). This indicates that these receptors have an established, proven history as therapeutic targets.

In consequence, there is a continuing medical need for identification and characterization of further receptors that can play a role in diagnosis, preventing, ameliorating or correcting dysfunctions, disorders, and diseases. Included among such diseases are a broad range of psychiatric and CNS disorders, including learning and memory dysfunctions.

**Summary of the Invention**

The G protein-coupled receptor of the present invention, IGPcR17, is especially useful for diagnosis, prevention, amelioration or correction of diseases associated with signal processing in the central nervous system (CNS). In particular, IGPcR17 satisfies a need in the art for identification and characterization of further receptors that can play an important role in diagnosis, prevention, amelioration or correction of psychiatric and central nervous system diseases and disorders, especially learning and memory dysfunctions, disorders or diseases, such as aphasia and Alzheimer's disease; movement dysfunctions, disorders or diseases, such as tics, tremor, Tourette's syndrome, Parkinson's disease, Huntington's disease, dyskinesias, dystonia, pain and spasms.

The GPcR of the present invention, IGPcR17, is especially useful for diagnosis, preventing, ameliorating or correcting of diseases associated with signal processing in the CNS, particularly diseases relating to memory processing and failure: *i.e.* learning and memory dysfunctions, including long term and short term memory disorders. Modulation of the activity of IGPcR17 may also be used for purposes of memory enhancement.

Embodiments of the invention include an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO:1, or any unique fragment thereof, particularly wherein the nucleotide sequence of the fragment is greater than ten base pairs in length. Embodiments also include an isolated polynucleotide which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or any unique fragment thereof, particularly wherein the amino acid sequence of the fragment is greater than ten amino acids in length. Embodiments of the invention include any isolated nucleic acid molecule or polynucleotide comprising an allelic variant of a nucleotide sequence or polynucleotide which encodes a polypeptide comprising the amino acid sequence of

SEQ ID NO:2, wherein said allelic variant retains at least 70% nucleic acid homology, or in increasing preference at least 80%, 85%, 90%, 95% or 98% nucleic acid homology and hybridizes to the complement of SEQ ID NO:1 under stringent conditions (Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York): also included are such isolated nucleic acid molecules or polynucleotides that comprise a nucleotide sequence which encodes at least one of the group of polypeptides, peptides and fusion proteins, comprising an amino acid sequence at least 70% similar, or in increasing preference at least 75%, 80%, 85%, 90%, 95% or 98% similar, to SEQ ID NO:2.

Vectors comprising an isolated nucleic acid molecule or polynucleotide of the invention as previously described are a further embodiment of the invention.

Additional embodiments include host cells genetically engineered to contain such a vector or genetically engineered to contain such a nucleic acid molecule or polynucleotide of the invention as described above, and particularly wherein the nucleic acid molecule or polynucleotide of the invention is operatively linked with a nucleotide regulatory sequence that controls expression of said nucleic acid molecule or polynucleotide in the host cell. Also included are host cells which are drawn from prokaryotic bacterial cells, or from eukaryotic cells, particularly or yeast, insect or mammalian cells, preferred embodiments employing a mammalian host cell being those in which the host cell is a CHO, BHK, COS, CV1, 293, fibroblast or VERO cell. Embryonic stem cells containing a disrupted endogenous IGPcR17 gene are also preferred embodiments of the invention, the most preferred embryonic stem cells being derived from mice.

Preferred embodiments of the invention include antibodies to the IGPcR17 protein, polypeptides, peptides, isolated domains and fusion proteins.

Agonists and antagonists of IGPcR17 are preferred embodiments of the invention,

including: (a) 'small molecules' of molecular mass less than 6 kDa; (b) molecules of intermediate size, having molecular mass between 5 kDa to 15 kDa; and (c) large molecules of molecular mass greater than 12 kDa; the latter including mutant natural IGPcR17 ligand proteins that compete with native natural IGPcR17 ligand and which modulate IGPcR17 gene expression or gene product activity. Preferred embodiments of the invention are those wherein such molecules bind specifically to the IGPcR17 receptor or to the IGPcR17 gene. Further embodiments are methods of identifying such compounds which modulate the activity of the IGPcR17 receptor or of IGPcR17 gene expression, such as anti-sense and ribozyme molecules that can be used to inhibit IGPcR17 gene expression, or expression constructs that are capable of enhancing IGPcR17 gene expression.

The non-human animal orthologs of the human sequence in SEQ ID NO:1 are preferred embodiments of the invention, particularly ungulate and rodent sequences, and especially those of rat and mouse, and also polynucleotides comprising these sequences or homologous or partially homologous sequences as indicated for the human nucleic acid and polynucleotide. Preferred embodiments include polynucleotides of such non-human animal orthologs comprising a nucleotide sequence which encodes a polypeptide comprising an amino acid sequence at least 70% similar, or in increasing preference at least 75%, 80%, 85%, 90%, 95% or 98% similar, to SEQ ID NO:2; or being at least ten amino acid residues in length and bearing the stated similarity to a unique part of SEQ ID NO:2.

Embodiments of the invention include knock-out animals which are non-human animals and which do not express IGPcR17. Preferred embodiments are those wherein the endogenous animal ortholog is functionally disrupted by homologous recombination methods such as conditional knock-out and/or null allele knock-out of the IGPcR17 gene. Mutated animals that express a non-functional or partially functional form of IGPcR17 are further embodiments of the invention. Embodiments of the invention also include progeny of the non-human animals described as being embodiments of the invention, the term 'progeny' including both

heterozygous and homozygous offspring. Further embodiments are non-human transgenic animal models expressing the human IGPcR17 cDNA sequence as shown in SEQ ID NO:1 or a modification thereof as described above, operatively linked to a nucleotide regulatory sequence that controls expression of the nucleic acid molecule in the host animal. Particularly preferred embodiments are those non-human animals (also termed animal models) in which the human IGPcR17 is encoded by a nucleic acid sequence which is homozygous in the animal model. In each embodiment of the invention comprising a non-human animal, preferable embodiments are those wherein the non-human animal is a mammal, particularly ungulate or rodent, and preferably wherein the non-human animal is from a genus selected from the group consisting of *Mus* (*e.g.*, mice), *Rattus* (*e.g.*, rats), *Oryctologus* (*e.g.*, rabbits) and *Mesocricetus* (*e.g.*, hamsters), mouse being the most preferable of this group.

Embodiments of the invention include primary cells and cell lines derived from any of the non-human animals of the invention, particularly the non-human transgenic animal models of the invention. Further embodiments include the amino acid sequence of those non-human animal orthologs of IGPcR17 that comprise an amino acid sequence at least 70% similar, or in increasing preference at least 75%, 80%, 85%, 90%, 95% or 98% similar, to the sequence of the mouse ortholog provided (SEQ ID NO:8); or a part of said non-human animal sequence which is at least ten amino acid residues in length and bears the stated similarity to a unique part of SEQ ID NO:8.

The use of the non-human animal or animal model of the invention, as described above, for the dissection of the molecular mechanisms of the IGPcR17 pathway, for the identification and cloning of genes able to modify, reduce or inhibit the phenotype associated with IGPcR17 activity or deficiency, constitutes a further embodiment of the invention, as does the use of such non-human animal or animal model for the identification of gene and protein diagnostic markers for diseases, for the identification and testing of compounds useful in the prevention or treatment of symptoms associated with IGPcR17 activity or deficiency, in particular but not

limited to central nervous system disorders, including neurologic, psychiatric and behavioral disorders, metabolic disorders, visual and olfactory disorders, and especially in the case of IGPcR17, visual diseases associated with signal processing in the CNS, particularly diseases relating to memory processing and failure: *i.e.* learning and memory dysfunctions, including long term and short term memory disorders.

Additional embodiments of the invention include methods of identifying compounds suitable for modulating the activity of the protein or polypeptide of the invention, as described above, for treatment of diseases characterized by aberrant expression or activity of IGPcR17. Preferred embodiments include methods of prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17, by the administration of compounds that bind specifically to the IGPcR17 gene or protein and/or which modulate IGPcR17 expression or IGPcR17 activity; the compounds that that bind specifically to the IGPcR17 gene or protein and/or which modulate IGPcR17 expression or IGPcR17 activity for the prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17; and the use of compounds that that bind specifically to the IGPcR17 gene or protein and/or which modulate IGPcR17 expression or IGPcR17 activity for prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17. Further preferred embodiments are gene therapy methods of prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17, by the administration of vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7, that modulate IGPcR17 expression or IGPcR17 activity; the vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7 which modulate IGPcR17 expression or IGPcR17 activity for the prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17; and the use of vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7 which modulate IGPcR17 expression or IGPcR17 activity for prevention, amelioration or treatment of diseases characterized by aberrant

expression or activity of IGPcR17.

## Brief Description of the Figures

Figure 1: Fig. 1 depicts the full-length coding DNA (cDNA) sequence of the human IGPcR17 gene (SEQ ID NO:1).

Figure 2: Fig. 2 depicts the amino acid sequence of the human IGPcR17 protein (SEQ ID NO:2).

Figure 3: Fig. 3 depicts Northern blots detecting human IGPcR17 transcript in RNA from several human tissues.

Figure 4: Fig. 4a depicts an autoradiogram of a Multi Tissue Expression Array dot blot membrane (described in Fig. 4b) hybridized with a human IGPcR17 probe;

Fig. 4b depicts the diagram of a Multiple Tissue Expression Array membrane (Clontech Laboratories, Palo Alto, CA; cat no. 7775-1) indicating type and position of human poly A+ RNAs dotted onto a nylon membrane.

Figure 5: Fig. 5 depicts the full-length coding DNA (cDNA) sequence of the mouse IGPcR17 gene (SEQ ID NO:8).

Figure 6: Fig. 6 depicts the amino acid sequence of the mouse IGPcR17 protein (SEQ ID NO:9).

Figure 7: Fig. 7a depicts a comparison of the amino acid sequence of the human IGPcR17 to the amino acid sequence of the mouse ortholog protein of human IGPcR17;

Fig. 7b depicts a comparison of the amino acid sequence of the human IGPcR17 to the amino acid sequence of the human KIAA0001 protein.

Figure 8: Fig. 8 depicts hydropathy plots comparing the human IGPcR17 receptor, the human KIAA0001 7TM receptor and the mouse ortholog protein of the human IGPcR17 receptor.

Figure 9: Fig. 9 depicts Northern blots detecting mouse IGPcR17 transcript in RNA from several mouse tissues.

Figure 10: Fig. 10 schematically outlines the construction of a mouse IGPcR17 targeting vector based on the method described by Wattler S & Nehls M, German patent application DE 100 16 523.0, "Klonierungssystem zur Konstruktion von homologen Rekombiationsvektoren", filed April 03, 2000, the major aspects of which are incorporated as Example 10.

## Detailed Description of the Invention

The present invention relates to the discovery, identification and characterization of nucleic acids that encode IGPcR17, a novel G protein-coupled receptor protein that contains regions of homology to the human KIAA0001 receptor. The invention encompasses nucleotide sequences encoding mammalian forms of IGPcR17, including human IGPcR17, nucleotides that encode some or all of its functional domains, such as extracellular domains (ECDs), the transmembrane domains (TMs), and the cytoplasmic domains (CDs); mutants of the IGPcR17 sequences, and fusion proteins of IGPcR17. The invention also encompasses host cell expression systems expressing such nucleotides, the host cells and expression products. The invention further encompasses IGPcR17 proteins, fusion proteins, antibodies to the receptor, antagonists and agonists of the receptor, transgenic animals that express an IGPcR17 transgene, recombinant knock-out animals that do not express the IGPcR17, and

animal models in which the IGPcR17 gene is mutated. The invention also encompasses compounds that modulate IGPcR17 gene expression or IGPcR17 receptor activity that can be used for drug screening, or for diagnosis, monitoring, preventing or treating visual dysfunctions associated with signal processing in the occipital lobe of the brain.

The invention further encompasses the use of IGPcR17 nucleotides, IGPcR17 proteins and peptides, as well as antibodies to IGPcR17, antagonists that inhibit ligand binding, receptor activity or expression, or agonists that increase ligand binding, activate receptor activity, or increase its expression, for the diagnosis and treatment of disorders, including, but not limited to treatment of central nervous system disorders. In addition, IGPcR17 nucleotides and proteins are useful for the diagnosis of an IGPcR17 or pathway abnormality, and for the identification of compounds effective in the treatment of disorders based on the aberrant expression or activity of IGPcR17. The invention also relates to host cells and animals genetically engineered to express the human IGPcR17 (or mutants thereof) or to inhibit or knock-out expression of the animal's endogenous IGPcR17 gene.

IGPcR17, as a new G protein-coupled receptor, can play a role in diagnosis, preventing, ameliorating and correcting diseases. These diseases include, but are not limited to, psychiatric and CNS disorders, including learning and memory dysfunctions, schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (COD), post traumatic stress disorders (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, multiple sclerosis, Alzheimer disease/dementia and other neurodegenerative diseases, severe mental retardation, dyskinesias, Huntington's disease, Gille de la Tourette's syndrome, tics, tremor, dystonia, spasms, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorders, epilepsy, migraine, attention deficit/hyperactivity disorder (ADHD), cardiovascular diseases, angina pectoris, including heart failure, angina pectoris, arrythmias, myocardial infarction, cardiac hypertrophy, hypertension, thrombosis,

arteriosclerosis, cerebral vasospasm, subarachnoid hemorrhage, cerebral ischenia, thrombosis, arteriosclerosis, peripheral vascular disease, Raynaud's disease, kidney disease – *e.g.* renal failure; dyslipidemias, obesity, emesis, gastrointestinal disorders, including irritable bowel syndrome (IBS), inflammatory bowel syndrome (IBD), diarrhoea, gastresophagal reflux disease (GERD), motility disorders and conditions of delayed gastric emptying, such as post operative or diabetic gastroparesis, and diabetis ulcers; other diseases including osteoporosis; inflammations; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; chemotherapy induced injury; tumor invasion; immune disorders; autoimmune diseases; urinary retention; asthma, allergies; arthritis; benign prostatic hypertrophy; endotoxin shock; sepsis; complication of diabetis mellitus; and gynaecological disorders.

In particular the new GPcR IGPcR17 satisfies a need in the art for identification and characterization of further receptors that can play an important role in diagnosis, preventing, ameliorating or correcting of diseases associated with signal processing in the CNS, particularly diseases relating to memory processing and failure: *i.e.* learning and memory dysfunctions, including long term and short term memory disorders. Modulation of the activity of IGPcR17 may also be used for purposes of memory enhancement. Although the signal that astrocytes send to trigger synapse formation by neurons has not yet been identified, it is proposed that a G protein-coupled receptor system recognizing the astrocyte trigger would be capable of generating and accounting for the observed complexity and plasticity of this process. Moreover, the characteristics of the G protein-coupled receptor IGPcR17 disclosed herein are appropriate for performing such a function.

**Definitions**

As used herein, the following terms, whether used in the singular or plural, have the meanings indicated.

IGPcR17 nucleotides, sequence or coding sequences – encompass DNA, including genomic DNA (e.g. the IGPcR17 gene), cDNA, RNA and include nucleotide sequences encoding IGPcR17 protein, peptide fragments, or fusion proteins.

IGPcR17 – means natural, or mature, IGPcR17 receptor protein. Polypeptides or peptide fragments of IGPcR17 protein are referred to as IGPcR17 polypeptides or IGPcR17 peptides. Fusions of IGPcR17, or IGPcR17 polypeptides or peptide fragments to an unrelated protein are referred to herein as IGPcR17 fusion proteins. ECD – means "extracellular domain" of the receptor protein; TM – means "transmembrane domain" and CD – means "cytoplasmic domain". A functional IGPcR17 refers to a protein which binds natural IGPcR17 ligand with high affinity and specificity *in vivo* or *in vitro*.

Ligand – a molecule that selectively binds to a receptor.

Receptor – a plasma membrane protein which binds one or more appropriate ligands and propagates their regulatory signals to target cells, either by direct intracellular effects, or by promoting the synthesis and/or release of another regulatory molecule known as a second messenger.

Agonist – a molecule, being a ligand and/or drug, that acts on one or more physiological receptors and mimics the effects of the endogenous regulatory compounds; generally these are compounds that activate the receptor.

Antagonist – a molecule, being a ligand and/or drug that inhibits a receptor, most acting by inhibiting the action of an agonist, for example by competing for agonist binding sites on a receptor. These are generally themselves devoid of intrinsic regulatory activity, but act to block receptor activation.

Transgenic animal – a non-human animal containing one or more additional, often

foreign genes or "transgenes", integrated into its genome, that can be used as model systems to determine the phenotypic effects of expressing those genes.

Knock-out or knock-out animal – a non-human animal wherein a transgene is inserted into the genome to create a partial or complete loss-of-function mutation of an endogenous gene. Endogenous genes are inactivated usually by homologous recombination, using replacement or insertion-type gene targeting vectors.

**Gene**

Novel GPcR genes may be isolated using expression cloning, by synthesizing specific oligonucleotides based on the sequence of purified proteins, using low stringency hybridization (Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York) and by degenerate PCR using known receptor sequences. GPcR genes may also be identified by large scale sequencing, as in the Human Genome Project, followed by analysis of expressed sequence tags (ESTs), or complete sequences present in databases. Known GPcR sequences or conserved regions thereof may be employed as query sequences to extract novel GPcR sequences from these databases.

The present invention provides IGPcR17, a novel G protein-coupled receptor protein described for the first time herein, and characterized as having seven hydrophobic domains which span the plasma membrane and which are connected by alternating extracellular and intracellular hydrophilic loops.

The IGPcR17 sequences encode a protein of 342 amino acids. At the amino acid level IGPcR17 has 47% identity and 66% sequence homology with conserved substitutions to human KIAA0001 sequences encoding a 338 amino acid protein (Genpept accession number NP055694, PID g7661848; Nomura *et al.*, 1994, DNA Res., 1:27-35; Nomura *et al.*, 1994, DNA Res., 1:47-56). KIAA0001 is related to

proteinase-activated receptors 1, 2 and 3, and to platelet activating factor receptor. It is widely expressed, with expression being highest in placenta, adipose tissue, stomach and intestine, moderate in brain, spleen, lung and heart, and low in kidney (Chambers *et al.*, 2000, J. Biol. Chem., 275:10767-10771). Recently, ligands to this orphan receptor were found: UDP-glucose and the closely related UDP-galactose, UDP-glucuronic acid and UDP-N-acetylglucoseamine, but not uridine or uridine nucleotides were found to activate recombinant KIAA0001 receptors. KIAA0001 was also shown to be coupled to G proteins of the Gi/o class (Chambers *et al.*, 2000, J. Biol. Chem., 275:10767-10771). The KIAA0001 gene is localized within the chromosome 3q region linked to Usher syndrome type III, representing progressive hearing loss and retinal degradation. However, no mutations were detected in the coding region among affected patients, although they might have been present in the gene regulatory regions (Joensuu *et al.*, 2000, Genomics, 63:409-416).

The mouse ortholog gene to the human IGPcR17 gene was identified by a search of public databases EMBL and Genbank using BLAST (Basic Local Alignment Search Tool, National Institutes of Health, Bethesda MD, U.S.A.) using the full-length coding sequence of the human gene as a query sequence. The highest score was assigned to an EMBL entry, EST/MM1179620: Soares mouse 3NbMs *Mus musculus* cDNA clone IMAGE 764455; length 542 bp. The EST open reading frame was used in primer design. From mouse strain 129, genomic clones containing the mouse IGPcR17 full-length cDNA (SEQ ID NO:6) and flanking genomic sequences were isolated and sequenced. Based on the completed full-length cDNA sequence, the predicted mouse IGPcR17 amino acid sequence (SEQ ID NO:7) revealed that the amino acid identity of human IGPcR17 to mouse IGPcR17 is 89 %.

In particular, the invention encompasses sequences coding for IGPcR17 polypeptides, or functional domains of the IGPcR17, mutated, truncated or deleted IGPcR17, and IGPcR17 fusion proteins. The invention also encompasses nucleotide constructs that inhibit expression of the IGPcR17 gene, such as anti-sense and ribozyme constructs, or enhance expression of IGPcR17 in combination with

regulatory sequences such as promoters and enhancers.

The cDNA sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) of human IGPcR17 of this invention are shown in Fig. 1 and Fig. 2. The IGPcR17 nucleotide sequences of the invention include the DNA sequence shown in Fig. 1, nucleotide sequences that encode the amino acid sequence shown in Fig. 2 and any nucleotide sequence that hybridizes to the complement of the DNA sequence shown in Fig. 1 under highly stringent conditions (Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York). Functional equivalents of the IGPcR17 gene product include naturally occurring IGPcR17, mutant and degenerate variants present in humans and other species. Preferred IGPcR17 nucleic acids encode polypeptides that are at least 55% identical or similar to the amino acid sequence shown in Fig. 2. Nucleic acids which encode polypeptides which are at least 70%, and even more preferably, in increasing order of preference, at least 80%, 85%, 90%, 95%, or 98% identical or similar. In a particularly preferred embodiment, the nucleic acid of the present invention encodes a polypeptide having an overall amino acid sequence homology or identity of, in increasing order of preference, at least 70%, 80%, 85%, 90%, 95%, 98%, or at least 99% with the amino acid sequence shown in Fig. 2.

The invention also provides DNA molecules that are the complements of the nucleotide sequences described above and which may act as IGPCR17 anti-sense molecules useful in IGPcR17 gene regulation. Orthologs of the human IGPCR17 gene present in other species can be identified and readily isolated. They can be useful for developing cell and animal model systems for purposes of drug discovery. For example, cDNA or genomic DNA libraries derived from the organism of interest can be screened by hybridization using the nucleotides described above, or by performing PCR using degenerate oligonucleotide primers. (See Sambrook *et al.*, 1989, "*Molecular Cloning, A Laboratory Manual*", Cold Spring Harbor Press, New York, USA; and Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular

Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., New York). Additionally, expression libraries can be screened using standard antibody screening techniques or by doing database searches for homologues and then cloning them based on the sequence. The identified sequences may be sub-cloned and sequenced. The IGPcR17 gene sequences may additionally be used to isolate mutant IGPcR17 gene alleles, or to detect defects in the regulatory sequences of the IGPcR17 using DNA obtained from an individual suspected of or known to carry the mutant IGPcR17 allele. Mutant alleles may be isolated from individuals either known or proposed to have a genotype which contributes to the symptoms of disorders arising from the aberrant expression or activity of the IGPcR17 protein. The isolation of human genomic clones is helpful for designing diagnostic tests and therapeutics. For example, sequences derived from the human gene can be used to design primers for use in PCR assays to detect mutations for diagnostics.

The nucleotides of this invention are also preferred for use in mapping the location of the gene to the chromosome, in a process termed chromosomal mapping. Various techniques known to those skilled in the art, including but not limited to *in situ* hybridization of labeled probes to flow-sorted chromosomes, fluorescence *in situ* hybridization (FISH) and PCR mapping of somatic cell hybrids may be employed. This allows the physical location of gene regions to be associated with genetic diseases, based on a genetic map. Genetic linkage analysis can then be used to identify the relationship between genes and diseases (see Egeland *et al.*, 1987, Nature, 325:783-787). Preferred uses of this map include diagnostic tests and reagents, in pharmacogenetics studies and in monitoring patient responses to drugs in clinical trials.

**Proteins and polypeptides**

Fig. 2 shows the amino acid sequence of the human IGPcR17 protein. The amino acid sequence of IGPcR17 contains hydrophilic domains located between the transmembrane domains, arranging an alternating location of the hydrophilic

domains inside and outside the cell membrane. Polypeptides which are at least 70%, and even more preferably at least 80%, 85%, 90%, 95%, 98% or 99% identical or similar to the amino acid sequence represented by Fig. 2 are encompassed by this invention.

In particular, the invention encompasses IGPcR17 polypeptides, or functional domains of the IGPcR17, mutated, truncated or deleted IGPcR17, and host cell expression systems that can produce such IGPcR17 products. IGPcR17 proteins, polypeptides and peptides, can be prepared for the generation of antibodies, as reagents in diagnostic assays, in the identification of other cellular gene products involved in regulating IGPcR17, as reagents for screening for compounds that can be used in the treatment of conditions involving IGPcR17, and as pharmaceutical reagents useful in the treatment of related disorders.

The invention also encompasses proteins that are functionally equivalent to the IGPcR17 encoded by the nucleotide sequences, as defined by the ability to bind natural IGPcR17 ligand, the resulting biological effect of natural IGPcR17 ligand binding, *e.g.*, signal transduction, a change in cellular metabolism or change in phenotype. Such functionally equivalent IGPcR17 proteins include but are not limited to additions or substitutions of amino acid residues, which result in a silent change. Also preferred in this invention are mutant IGPcR17 proteins with increased function, and/or greater signaling capacity; or decreased function, and/or decreased signal transduction capacity which may be generated by random mutagenesis techniques and site-directed mutagenesis techniques well known to those skilled in the art. The same strategy can also be used to design mutant forms of IGPcR17 based on the alignment of human IGPcR17 and IGPcR17 orthologs from other species. Highly preferred are other mutations to the IGPcR17 coding sequence that can be made to generate IGPcR17 constructs that are better suited for expression, scale up, *etc.* in the host cells chosen. Host cells may be chosen depending on their varying capacity to modify synthesized proteins.

Peptides corresponding to one or more domains of the IGPcR17 (*e.g.*, ECD, TM or CD), truncated or deleted IGPcR17s, as well as fusion proteins are also within the scope of the invention and can be designed on the basis of the IGPcR17 nucleotide and IGPcR17 amino acid sequences disclosed above. Such IGPcR17 polypeptides, peptides and fusion proteins can be produced using techniques well known in the art for expressing protein encoding IGPcR17 sequences. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See Sambrook *et al.*, 1989, "*Molecular Cloning, A Laboratory Manual*", Cold Spring Harbor Press, N.Y.; and Ausubel FM *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., New York.). A variety of host-expression vector systems may be utilized to express the IGPcR17 nucleotide sequences of the invention. The IGPcR17 peptide or polypeptide may be anchored in the cell membrane and purified or enriched from such expression systems using appropriate detergents and lipid micelles, and methods well known to those skilled in the art. Or, where the IGPcR17 peptide or polypeptide is secreted by the cells, it may be isolated from the culture media. Such host cells themselves may be used to assess biological activity, *e.g.*, in drug screening assays.

The expression systems that may be used for purposes of the invention include, but are not limited to microorganisms such as bacteria (*e.g.*, *E. coli, B. subtilis*); yeast (*e.g.*, *Saccharomyces sp., Pichia sp.*); insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus); plant cell systems infected with recombinant viral or plasmid expression vectors; or mammalian cell systems (*e.g.*, COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing mammalian promoters. Lower amounts of functional protein are expressible in *E. coli* and yeast, particularly as *E. coli* do not contain G proteins or effectors. G proteins may be added to *E.coli* expressing G protein-coupled receptors in cell membrane, in the cell-based assays. Yeast cells may be humanized by co-transfixing human G proteins. The yeast *Pichia pastoris* is preferred over *Saccharomyces cerevisiae* for purification of G protein-coupled receptors for

structural studies. The most preferred systems for expression are the baculovirus/insect cell and mammalian cell systems, as they can produce the largest quantities of G protein-coupled receptors in functional form for analysis. Mammalian cells are preferred because they express the necessary G proteins, and vaccinia and Semliki Forest virus are preferred as vectors. (See Tate *et al.*, 1996, Tibtech 14:426-430).

**Diagnostic and therapeutic reagents and kits**

In one embodiment of the invention, the invention encompasses antibodies directed against IGPcR17 proteins or peptides, or IGPcR17 fusion proteins, as described above. Such antibodies include but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, anti-idiotypic (anti-Id) antibodies, including Fab fragments. The antibodies may be generated and purified, or conjugated according to methods well known in the art. See for example Harlow E and Lane D, 1988, "*Antibodies: A Laboratory Manual*", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, which is incorporated herein by reference in its entirety.

In another embodiment, the antibodies of the invention may be used, for example, as part of a diagnostic or a prognostic, and as a part of compound screening schemes, for the evaluation of the effect of test compounds on expression and/or activity of the IGPcR17 gene product. Preferably, antibodies may be used in therapeutic regimes as a method for the inhibition of abnormal IGPcR17 activity. Also preferred are antibodies directed against wild type or mutant IGPcR17 gene products or conserved variants or peptide fragments thereof to detect the pattern and level of expression, as well as distribution in tissues, of the IGPcR17 in the body, also by *in situ* detection. The antibodies may be employed as part of an enzyme immunoassay (EIA), a radioimmunoassay, or as an antibody labeled with a chemiluminescent or a fluorescent compound.

In yet another embodiment of the invention, the IGPcR17 proteins or peptides, IGPcR17 fusion proteins, IGPcR17 nucleotide sequences, antibodies, antagonists and agonists can be useful for the detection of mutant forms of IGPcR17 or inappropriately expressed forms of IGPcR17, for the diagnosis of disorders including but not limited to central nervous system disorders, neurologic, psychiatric and behavioral disorders, metabolic disorders, visual and olfactory disorders, immune, neuroimmune, neuroendocrine and inflammatory disorders and diseases. DNA encoding IGPcR17 or parts thereof may be used in hybridization or amplification assays of biological samples to detect abnormalities involving IGPcR17 gene structure, including point mutations, insertions, deletions and chromosomal rearrangements. Such genotyping assays may include, but are not limited to Southern analyses, single stranded conformational polymorphism analyses (SSCP), and PCR analyses (See Mullis KB, U.S. Pat. No. 4,683,202), the use of restriction fragment length polymorphisms (RFLPs), of variable numbers of short, tandemly repeated DNA sequences between the restriction enzyme sites (see Weber, U.S. Pat. No. 5,075,217), and by detecting and measuring IGPcR17 transcription.

Also within the scope of the invention are the IGPcR17 proteins or peptides, IGPcR17 fusion proteins, IGPcR17 nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and animals. These can be used for screening for drugs effective in the treatment of disorders. The use of engineered host cells and/or animals may offer an advantage in that both compounds that bind to the ECD of the IGPcR17 and compounds that affect the signal transduced by the activated IGPcR17 may be identified.

**Screening for receptor modulating agents**

In another embodiment of the invention, the invention encompasses the pharmacological testing wherein the cloned IGPcR17 genes are expressed in yeast, insect or mammalian cells and screened for a response to cognate or surrogate agonists. The agonists may be present in, but are not limited to, biological extracts,

peptide libraries and/or complex compound collections. The invention provides for screening which may utilize libraries of known compounds, including natural products or synthetic chemicals, and biologically active materials, including proteins, for compounds which are inhibitors or activators. Candidate test compounds include all kinds of combinatorial chemistry derived molecular libraries of amino acids, peptides, soluble peptides, modified peptides, antibodies, small organic and inorganic molecules.

In a further embodiment of the invention, a labeled test compound can be incubated with the receptor to determine whether one binds to the other. Functional assays including fibroblast and BM transformation assays, cell cycle analysis can be performed; as well as responses using signal transduction assays, including protein phosphorylation, guanylate cyclase activity, ion fluxes (*e.g.* calcium) and pH changes can be measured. High throughput drug screening systems are most preferred and may use assays including, but not limited to, the production of intracellular second messengers, such as cAMP, diacylglycerol and inositol phosphates; the activation of reporter gene transcription, such as luciferase and beta-galactosidase under for example the cAMP-responsive element; receptor-mediated actions on adenylyl cyclase and phospholipase C leading also for example to dispersion or aggregation of frog melanophores. (Reviewed in Tate *et al.*, 1996, Tibtech 14:426-430; included in entirety herein).

In a highly preferred embodiment, a functional genomics approach for protein-protein interaction screening may be employed wherein the GPcR is produced in "humanized yeast cells": expression in yeast along with endogenous or promiscuous mammalian or human G-alpha proteins. Transient expression of cDNA can also be carried out using mammalian CHO, HEK-293 cells or COS-7 cells and receptors can be analyzed for ligand binding and drug interactions (for example as described in Fraser *et al.*, 1995, J. Nucl. Med., 36:17S-21S). Also preferred is site-directed mutagenesis to define regions of IGPcR17 that have functional importance. Site-directed mutagenesis may be used to map ligand-binding pockets and to identify

residues important for receptor interaction and activation. Compounds that can be generated using modeling methods to bind these residues are also within the scope of this invention. For example, receptor down-regulation and the development of drug tolerance, such as seen in asthma patients who use bronchial dilators which are beta-adrenergic agonists leading to tachyphylaxis, can be studied in these cell systems. The expression of both intact and hybrid receptors is preferred. The number of receptors, as well as mRNA levels can be measured. Agents for radionuclide imaging to monitor level changes can be developed.

Some of the known receptors and their ligands defined by above techniques are shown below.

**Ligand screening**

| Ligand Categories | Examples of Ligands | Examples of Receptors |
|---|---|---|
| Peptides | Angiotensin, bradykinin, EGF, NPY, neurokinins, PAF, ACTH, C5a, IL8 | Leukocyte receptor |
| Steroids | Testosterone, progesterone, FSH, TSH | Testosterone-R, progesterone-R, FSH-R, TSH-R |
| Prostaglandins | thromboxane | thromboxane receptor |
| Neuro-transmitters | Adenosine, adrenergic, dopamine, muscarinic, purinergic, serotonin, opioid | adrenergic receptors, purinergic P2U-R, P2Y1-R |
| Second messengers | ATP, UDP, cAMP, U, T, adenylyl cyclase, inositol phosphate, | P2X-R, P2Y-R |
| Ion channels | Calcium, sodium, potassium, chloride | Ligand-gated ion channels |
| Regulatory sites | Benzodiazepines, glycine, MK-801 | Glycine receptors |

| | | |
|---|---|---|
| Uptake sites | Adenosines, choline, dopamine, GABA, glutamate | Dopaminergic receptors, GABA, glutamate receptors |
| Nucleotide sugars | UDP-glucose, ADP-ribose | KIAA0001 |

The invention encompasses antagonists and agonists of IGPcR17, as well as compounds or nucleotide constructs that inhibit expression of the IGPcR17 gene (anti-sense and ribozyme molecules), or promote expression of IGPcR17 (wherein IGPcR17 coding sequences are operatively associated with promoters, enhancers, *etc.*). Highly preferred are the IGPcR17 protein products (especially soluble derivatives of IGPcR17, or truncated polypeptides lacking the TM or CD domains) and fusion protein products, antibodies and anti-idiotypic antibodies, antagonists or agonists (including compounds that modulate signal transduction which may act on downstream targets in the IGPcR17 signal transduction pathway) that can be used for therapy of such diseases, by inhibiting receptor activity.

Nucleotide constructs encoding functional forms of IGPcR17 and mutant forms of IGPcR17 are preferred embodiments of the invention, as their uses include employment in the genetic engineering of host cells. Other preferred embodiments of the invention are anti-sense and ribozyme molecules, preferred for use in "gene therapy" approaches in the treatment of disorders or diseases arising from the aberrant or altered activity of IGPcR17. The gene therapy vector alone or when incorporated into recombinant cells, may be administered in a suitable formulation for intravenous, intra-muscular, intra-peritoneal delivery, or may be incorporated into a timed release delivery matrix.

**Transgenic and knock-out animal models**

The animal-based and cell-based models can be used to identify drugs, biologicals, therapies and interventions which can be effective in treating disorders with aberrant

expression or activity. IGPcR17 sequences can be introduced into, and over-expressed and/or can be disrupted in order to under-express or inactivate IGPcR17 gene expression.

In one embodiment of the invention, the IGPcR17 gene products can also be expressed in transgenic animals. Non-human animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, sheep, cows, goats, may be used to generate IGPcR17 transgenic animals. The present invention provides for transgenic animals that carry the IGPcR17 transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals. The transgene may be expressed in all tissues of the animal, or may be limited to specific tissues. Any technique known in the art may be used to introduce the IGPcR17 transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe PC and Wagner TE, U.S. Patent No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten *et al.*, 1985, Proc. Natl. Acad. Sci., USA 82:6148-6152); gene targeting in embryonic stem cells (Thompson *et al.*, 1989, Cell 56:313-321); electroporation of embryos (Lo, 1983, Mol. Cell. Biol., 3:1803-1814); and sperm-mediated gene transfer (Lavitrano *et al.*, 1989, Cell 57:717-723); *etc.* For a review of such techniques, see Gordon, 1989, "*Transgenic Animals*", Intl. Rev. Cytol. 115:171-229, which is incorporated by reference herein in its entirety.

The present invention relates to knock-out animals engineered by homologous recombination to be deficient in the production of the IGPcR17. The present invention is directed to a knock-out animal having a phenotype characterized by the substantial absence of IGPcR17, otherwise naturally occurring in the animal. In addition, the invention encompasses the DNA constructs and embryonic stem cells used to develop the knock-out animals and assays which utilize either the animals or tissues derived from the animals. Preferably, these cells, tissues and cell lines are characterized by the substantial absence of IGPcR17 that would otherwise be naturally occurring in their normal counterparts.

Gene targeting is a procedure in which foreign DNA sequences are introduced into a specific locus within the genome of a host cell. In another embodiment of the invention, endogenous IGPcR17 gene expression can be reduced by inactivating or knocking out the IGPcR17 gene or its promoter using targeted homologous recombination. (*e.g.*, see Smithies *et al.*, 1985, Nature 317:230-234; Thomas & Capecchi, 1987, Cell 51:503-512; Thompson *et al.*, 1989, Cell 5:313-321; each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional IGPcR17 (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous IGPcR17 gene (either the coding regions or regulatory regions of the IGPcR17 gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express IGPcR17 *in vivo*. Insertion of the DNA construct, via targeted homologous recombination into the genome, results in abolishing IGPcR17 gene function.

One preferred technique for targeted mutagenesis in this invention is based on homologous recombination. The general methodologies of targeting mutations into the genome of cells, and the process of generating mouse lines from genetically altered embryonic stem (ES) cells with specific genetic lesions are well known (Bradley, 1991, Cur. Opin. Biotech. 2: 823-829). See also U.S. patents 5,557,032 by Mak *et al.*, and U.S. Patent No. 5,487,992 by Capecchi *et al.*, included by reference herein. Preferred in this invention is a synthetic recombination vector which contains the genetic information of the targeted chromosomal locus recombines with the genomic DNA after introduction into a cell. A strategy of "positive/negative selection" can be used to enrich the cell population for cells in which targeting vectors have integrated into the host cell genome, and recombination has occurred at the desired gene locus (Mansour, *et al.*, 1988, Nature 336:348). The vector usually contains a positive selection cassette which is flanked by the genetic information of the target locus to enrich for cells where the vector successfully recombines with the chromosomal DNA against the pool of non-recombinant cells.

The likelihood of obtaining an homologous recombination event increases with the size of the chromosomal vector DNA and is further dependent on the isogenicity between the genomic DNA of the vector and the target cell (See te Reile *et al.*, 1992, P.N.A.S. USA 89:5128-5132; Deng *et al.*, 1991, Mol. Cell. Biol., 12, 3365-3371). Also preferred in this invention are large stretches of genomic DNA flanking the IGPcR17 gene ortholog in the target animal species. The cloning of large chromosomal fragments of the target gene, the sub-cloning of this DNA into a bacterial plasmid vector, the mapping of the gene structure, the integration of the positive selection cassette into the vector and finally, the flanking of one or both homologous vector arms by a negative selection marker are well described in the literature. Also preferred are replacement-type targeting vectors using yeast host cells are described by Storck *et al.*, 1996, Nuc. Acids Res. 24:4594-4596. The use of other vectors such as bacteriophage λ and vectors for phage-plasmid recombination have been described by Tsuzuki *et al.*, 1998, Nuc. Acids Res 26:988-993; transposon-generated gene targeting constructs have also been described by Westphal *et al.*, 1997, Curr. Biol., 7:530-533 and are within the scope of the invention.

The most highly preferred method in this invention is described by Wattler S & Nehls M, German patent application DE 100 16 523.0, "Klonierungssystem zur Konstruktion von homologen Rekombiationsvektoren", filed April 03, 2000, included by reference in whole herein,. and described in part in Example 7. This method reduces the time required for the construction of such vectors from 3-6 months to about 14 days. The vector includes a linear lambda vector (lambda-KO-*Sfi*) that comprises a stuffer fragment; an *E. coli* origin of replication; an antibiotic resistance gene for bacterial selection, two negative selection markers suitable for use in mammalian cells; LoxP sequences for cre-recombinase mediated conversion of linear Lambda phages into high copy plasmids. In a final targeting vector, the stuffer fragment is replaced by nucleotide sequences representing a left arm of homology, an ES cell selection cassette, and a right arm of homology.

The transformation of mouse129 ES cells with the final vector construct is done according to standard procedures. The targeting vector is linearized and then introduced by electroporation into ES cells. Cell clones are positively selected with G418 and negatively selected with GANC (ganciclovir, 0.2 μM). Targeted ES-cell clones with single integration sites are identified, confirmed by hybridization, and expanded in culture for injection.

The invention also encompasses embryonic stem (ES) cells derived from a developing mouse embryo at the blastocyst stage, that are modified by homologous recombination to contain a mutant IGPcR17 gene allele. The modified ES cells are reintroduced into a blastocyst by microinjection, where they contribute to the formation of all tissues of the resultant chimeric animal, including the germ line (Capecchi, 1989, Trends Genet., 5:70; Bradley, *et al.*, 1984, Nature, 309:255). Modified ES cells may also be stored before reimplantation into blastocysts. The chimeric blastocysts are implanted into the uterus of a pseudopregnant animal, prepared by mating females with vasectomized males of the same species. Typically chimeras have genes coding for a coat color or another phenotypic marker that is different from the corresponding marker encoded by the stem cell genes.

Also within the scope of the invention are chimeric male non-human animals and their heterozygous offspring carrying the IGPcR17 gene mutation which are bred to obtain progeny which are homozygous for the mutation, preferred animals being mice. A phenotype selection strategy may be employed, or chromosomal DNA may be obtained from the tissue of offspring, screened using Southern blots and/or PCR amplification for the presence of a modified nucleotide sequence at the IGPcR17 gene locus, liked described in the above section of identifying positivlly targeted ES cells. Other means for identifying and characterizing transgenic knock-out animals are also available. For example, Northern blots can be used to probe mRNA obtained from tissues of offspring animals for the presence or absence of transcripts coding for either the IGPcR17, the marker gene, or both. In addition, Western blots might be used to assess IGPcR17 expression by probing with antibody specific for

the receptor.

These animals are characterized by including, but not limited to, a loss in the ability to bind ligands specific for IGPcR17 and/or by a loss in expression from the IGPcR17 gene locus. Preferably, the animals produce no functional forms of IGPcR17 at all. Once homozygous transgenic animals have been identified, they may preferably be interbred to provide a continual supply of animals that can be used in identifying pathologies dependent upon the absence of a functional IGPcR17 and in evaluating drugs in the assays described above. Also highly preferred in this invention, are these animals as providing a source of cells, tissues and cell lines that differ from the corresponding cells, tissues and cell lines from normal animals by the absence of fully functional forms of IGPcR17.

The methodology needed to make such animals can be adapted to any non-human animal, preferably rodents such as hamsters, rats or mice, and most preferably, mice. In another embodiment, clones of the non-human transgenic animals can be produced according to methods described in Wilmut *et al.*, 1997, Nature, 385:810-813.

## EXAMPLES

### Example 1. Identification of a full-length human cDNA coding for a novel GPcR, IGPcR17.

A coding sequence of 1026 bp was identified as follows from the EMBL alert HTGH (High Throughput Genome) database (see Fig.1, SEQ ID NO:1). A search was performed using the nucleotide sequences of known GPcRs. A sequence with a statistically significant score was returned and searched for open reading frames. Subsequently a putative coding region was assigned and used in primer design. The tracked human genomic IGPcR17 sequence contains the full-length cDNA sequence, identified as a single exon coding GPcR. IGPcR17 encodes a protein of 343 amino acids, (see Fig.2, SEQ ID NO:2).

A BLASTP search (Basic Local Alignment Search Tool for Proteins, National Institutes of Health, Bethesda MD, U.S.A.) revealed that the human protein most closely related to human IGPcR17 is KIAA0001, a human G protein-coupled receptor of 338 amino acids (Genpept. accession number NP055694; Nomura *et al.*, 1994, DNA Res. 1:27-35; Nomura *et al.*, 1994, DNA Res. 1:47-56). IGPcR17 has 47% identity and 66% sequence homology with conserved substitutions to KIAA0001.

**Example 2. Tissue-specific expression of human IGPcR17, analysis by RT-PCR.**

A panel of cDNAs derived from total RNA from 29 human tissues (Clontech Laboratories, Inc., Palo Alto CA, USA; Invitrogen Corp., Carlsbad CA, USA) was tested in a reverse transcription-polymerase chain reaction (RT-PCR) assay. The sequence of the primers used to amplify a 466 bp product (SEQ ID No:5), which spans a region including transmembrane domains 1 to 4, is as follows:

5'- CAC TGC TCT ACA CTG TCC TG (coding sequence position 86 to 105; SEQ ID NO:3)

5'- GAC CGA ACT CTG ATT TAA GG (coding sequence position 501 to 486; SEQ ID NO:4)

The conditions for the PCR were: denaturation at 94°C for 45 seconds, annealing at 56°C for 1 minute, and extension at 72°C for 30 seconds, for a total of 35 cycles, in a Thermocycler (MJ Research, Watertown MA, USA; type PTC-225). The PCR products were analyzed on an 1.8% agarose gel and stained with ethidium bromide to visualize DNA by ultraviolet imaging. The tissues analyzed were: skin, whole brain, fetal brain, cerebellum, thymus, esophagus, trachea, lung, breast, mammary gland, heart, liver, fetal liver, kidney, spleen, adrenal gland, pancreas, stomach, small intestine, skeletal muscle, adipose tissue, uterus, placenta, bladder, prostate, testis, colon, rectum and cervix. Positive (human genomic DNA) and negative (water)

controls were included.

Positive PCR products of 466 base pairs in size were observed in the human panel, in order of decreasing signal intensity, in whole brain, fetal brain, cerebellum, placenta, skin, adipose tissue, uterus, pancreas, spleen, fetal liver and trachea. The correct identity of the sequence amplified was confirmed by sequencing of the PCR products.

**Example 3. Tissue-specific expression of human IGPcR17 transcript, analysis by Northern hybridization.**

Northern hybridization of polyA+ RNAs from several human tissues was carried out using a IGPcR17 specific DNA-probe. The probe was generated by radiolabelling the purified and sequenced PCR product generated using primers as described in Example 2. The probe spans sequences coding for transmembrane regions 1 to 4 and is 466 bp in length. Commercially available Multiple Tissue Northern Blots (Clontech Laboratories, Palo Alto CA, USA) containing approximately 2 micrograms of poly A + RNA per lane, adjusted to provide a consistent beta-actin signal in each lane, were hybridized, following the manufacturer's instructions. These blots are optimized to give best resolution in the 1.0-4.0 kb range, and marker RNAs of 9.5, 7.5, 4.4, 2.4, 1.35 and 0.24 kb are run for reference. In addition, a different set of commercially available Multiple Tissue Northern Blots (BioChain Institute, Hayward CA, USA) containing 3 micrograms of human tissue poly A + RNA per lane, was hybridized. Membranes were pre-hybridized for 30 minutes and hybridized overnight at 68°C in ExpressHyb hybridization solution (Clontech Laboratories, Palo Alto CA, USA) as per the manufacturer's instructions. The cDNA probe used was labeled with [$\alpha^{32}$P] dCTP using a random primer labeling kit (Megaprime DNA labeling system, Amersham Pharmacia Biotech, Piscataway NJ, USA) and had a specific activity of 1 x $10^9$ dpm/µg. The blots were washed several times in 2X SSC, 0.05% SDS for 30-40 min at room temperature, and were then

washed in 0.1X SSC, 0.1% SDS for 40 min at 50°C (see Sambrook *et al.*, 1989, "*Molecular Cloning, A Laboratory Manual*", Cold Spring Harbor Press, New York, USA). The blots were covered with standard domestic plastic wrap and exposed to X-ray film at –70°C with two intensifying screens for 36 hours.

The tissues represented in the Clontech Laboratories Multiple Tissue Northern Blots are as follows:

| Blot #1 | Blot #2 |
|---|---|
| Heart | Stomach |
| Brain | Thyroid |
| Placenta | Spinal Cord |
| Lung | Lymph Node |
| Liver | Trachea |
| Skeletal Muscle | Adrenal Gland |
| Kidney | Bone Marrow |
| Pancreas | --- |

The results of this experiment indicate that IGPcR17 is expressed as an approximately 2.4 kb transcript in human brain & spinal cord (see Fig. 3a). A less intense signal was obtained at a lower molecular weight of about 1.8 kb in the trachea. This band is most probably a cross-reactive RNA species, as it is also observed upon hybridization with probes detecting a variety of GPcRs.

The tissues represented in the BioChain Institute Multiple Tissue Northern Blots are as follows:

| Blot #I | Blot #II | Blot #V | Blot#VIII |
|---|---|---|---|
| Heart | Stomach | Uterus | Brain |
| Brain | Jejunum | Cervix | Kidney |
| Liver | Ileum | Ovary | Spleen |

| | | | |
|---|---|---|---|
| Pancreas | Colon | Testis | Intestine |
| Skeletal Muscle | Lung | Prostate | Uterus |
| Lung | --- | Lung | Cervix |
| --- | --- | --- | Placenta |
| --- | --- | --- | Lung |

The results of this experiment indicate that IGPcR17 is expressed most strongly in a predominant 2.4 kb transcript in human brain. A fainter transcript species of about 4.4 kb in size is also detected in brain. Additional tissues expressing low levels of the IGPcR17 2.4 kb transcript are skeletal muscle, lung, intestine, placenta, uterus, cervix and prostate (see Fig. 3b).

**Example 4. Tissue-specific expression and relative abundance of human IGPcR17 transcript, dot blot analysis.**

Dot blot analysis of the IGPcR17 transcript in polyA+ RNAs from several human tissues was performed. An IGPcR17-specific DNA probe was generated by radiolabeling the purified and sequenced PCR product generated using primers as described in Example 2. The probe spans sequences coding for transmembrane domains 1 to 4 and is 466 bp in length. A commercially available Multiple Tissue Expression Array membrane (Clontech Laboratories, Inc., Palo Alto CA, USA, cat. no. 7775-1) containing polyA + RNA in each 1mm dot, normalized to the mRNA levels of eight housekeeping genes, was hybridized (see Fig. 4a and 4b). Thus changes seen can be attributed to actual differences in mRNA abundance. The membrane was pre-hybridized for 30 minutes and hybridized overnight, at 65°C in ExpressHyb Hybridization Solution as per the manufacturer's instructions (Clontech Laboratories, see above). The cDNA probe used was labeled with [$\alpha^{32}$P] dCTP using a random primer labeling kit (Megaprime DNA labeling system, Amersham Pharmacia Biotech, Piscataway NJ, USA) and had a specific activity of 1 x$10^9$ dpm/µg. The blot was washed several times in 2X SSC, 0.05% sodium dodecyl

sulfate (SDS) for 30-40 minutes at room temperature, then washed in 0.1X SSC, 0.1% SDS for 40 min at 50°C (see Sambrook *et al.*, 1989, "*Molecular Cloning, A Laboratory Manual*", Cold Spring Harbor Press, New York, USA). The blot was covered with plastic wrap and an X-ray film exposure made for two days at –70°C using two intensifying screens.

The results of this experiment indicate that IGPcR17 is expressed in all regions of the human brain and CNS examined, such as whole brain, cerebellum, substantia nigra, cerebral cortex, accumbens nucleus, frontal lobe, corpus callosum, thalamus, parietal lobe, amygdala, occipital lobe, caudate nucleus, spinal cord, temporal lobe, paracentral gyrus of cerebral cortex, medulla oblongata, pons and putamen, as well as, at a lower level in fetal brain (see Figure 4a).

**Example 5. Characterization of human IGPcR17 protein.**

Hydrophobic analysis of the predicted amino acid sequence showed seven hydrophobic regions corresponding to the seven transmembrane regions, a conserved structural feature of G protein-coupled receptors. Human IGPcR17 is a protein of 342 amino acid residues. The encoded protein was compared to amino acid sequences present in public databases EMBL and Genbank. IGPcR17 has 47% identity and 66% similarity to amino acid sequences of KIAA0001, a G protein-coupled receptor of 338 amino acids in size (Genpept accession number NP055694; Nomura *et al.*, 1994, DNA Res., 1:27-35; Nomura *et al.*, 1994, DNA Res., 1:47-56). Recently a rat ortholog of human KIAA0001, VTR 15-20 (Genpept accesion number O355881), encoding a protein of 305 amino acids was isolated and shown to have 81% amino acid identity, and 92% sequence homology (with conserved substitutions) to KIAA0001 (Charlton *et al.*, 1997, Brain Res., 764:141-148).

Fig. 7a shows the amino acid sequence of IGPcR17 ('query') compared to the amino acid sequence of the mouse ortholog of IGPcR17 ('sbjct'), and Fig. 7b shows the

amino acid sequence of IGPcR17 ('query') compared to the amino acid sequence of human KIAA0001 ('sbjct'); as abstracted from the SWISSPROT database and analyzed using a BLASTP alignment program. The predicted transmembrane domains of IGPcR17 are flanked by amino acids 29-50 (TM1), 61-80 (TM2), 100-124 (TM3), 142-160 (TM4), 190-212 (TM5), 238-257 (TM6), 282-306 (TM7), as underlined. Human IGPcR17 protein and the mouse ortholog IGPcR17 protein exhibit an amino acid identity of 88.7% in an overlap of 337 residues.

Fig. 8 shows a hydropathy plot for the predicted amino acid sequence of the human IGPcR17 protein compared to the sequences predicted for human KIAA0001 and the mouse ortholog IGPcR17. The analysis was performed using the method of Kyte and DooLittle (1982, J. Mol. Biol., 157:105-32), with the DAMBE program (Data Analysis in Molecular Biology and Evolution), University of Hong Kong, version 4.0.41.

**Example 6. Identification of mouse ortholog of human IGPcR17.**

The mouse ortholog gene to the human IGPcR17 gene was identified by a BLAST search of public databases EMBL and Genbank using the full-length coding sequence of the human gene as a query sequence. The highest score was assigned to an EMBL EST/MM1179620, Soares mouse 3NbMs *Mus musculus* cDNA clone IMAGE 764455; length 542 bp. The open reading frame was used in primer design. From mouse strain 129, genomic clones containing the full-length cDNA (SEQ ID NO:6, Fig. 5) and flanking genomic sequences were isolated and were utilized in the construction of targeting vectors. The cDNAs of the human and mouse IGPcR17 were found to be 90 % identical. Several clones were isolated from a mouse 129 genomic library by hybridization of a probe corresponding to a 356 bp PCR product, as described in Example 7.

**Example 7. Tissue-specific expression of mouse orthologue gene, analysis by RT-PCR.**

A panel of cDNAs derived from total RNA isolated from 22 mouse tissues was tested in a reverse transcription-polymerase chain reaction (RT-PCR) assay. The sequence of the primers used to amplify a 356 bp product (SEQ ID NO:10), which includes predicted transmembrane domains 3 to 5, is as follows:

5'- CGT TCC TGG GGT TGA TAA CC (coding sequence position 355- 374; SEQ ID NO:8)

5'- CTT GGG AAC TTT GGC TGA AC (coding sequence position 710-691; SEQ ID NO:9)

The conditions for the polymerase chain reaction (PCR) were: denaturation at 94°C for 45 seconds, annealing at 56°C for 1 minute and extension at 72°C for 30 seconds, for a total of 35 cycles, in a Thermocycler (MJ Research; type PTC-225). The PCR products were analyzed on an 1.8% agarose gel and stained with ethidium bromide to visualize DNA by ultraviolet imaging. The tissues analyzed were: brain, brain without cerebellum, cerebellum, thymus, eye, tongue, trachea, adipose tissue, lung, heart, spleen, kidney, liver, stomach, small intestine, large intestine, colon, rectum, bladder, uterus, prostate, testis. Positive (mouse genomic DNA) and negative (water) controls were included.

Positive PCR products of 356 bp in size were observed in the mouse panel in order of decreasing signal intensity in whole brain, brain without cerebellum, cerebellum, uterus, trachea, eye, bladder and rectum.

**Example 8. Tissue-specific expression of mouse IGPcR17, analysis by Northern hybridization.**

RNA originating from several tissues was isolated, separated by agarose gel

electrophoresis and blotted onto a Hybond N+ Nylon membrane (Amersham Pharmacia Biotech, Piscataway NJ, USA). After covalent fixation to the membrane by UV crosslinking, the Northern blot was hybridized with a mouse IGPcR-17 probe, as indicated in Example 7. Tissues analyzed are: total brain, cerebrum, cerebellum, heart, spleen, testis, epididymis and uterus.

A single transcript of about 2.4kb in size, was detected, being strongly expressed in total brain, cerebrum and cerebellum, and expressed in testis and epididymis to a weaker extent. In heart and spleen a single faint transcript species of about 2.8 kb in size was detected, as depicted in Figure 9.

**Example 9. Characterization of mouse IGPcR17 protein.**

Hydrophobic analysis of the predicted amino acid sequence of the mouse ortholog protein of IGPcR17 indicated seven hydrophobic domains corresponding to the seven transmembrane regions, a conserved structural feature of G protein-coupled receptors. The mouse IGPcR17 protein (Fig.6, SEQ ID NO:7) is a protein of 347 amino acid residues. The encoded protein has 88.7% identity to the amino acid sequence of the human IGPcR17 protein.

Figure 7a shows the amino acid sequence of human IGPcR17 ('query') compared to that of the mouse ortholog ('sbjct'), as abstracted from the SWISSPROT database and analyzed using the BLASTP alignment program. The predicted transmembrane domains (TM 1 to 7) of mouse IGPcR17 are flanked by amino acids 34-55 (TM1), 69-86 (TM2), 99-121 (TM3), 144-167 (TM4), 192-214 (TM5), 240-262 (TM6), 279-301 (TM7), as underlined.

Figure 8 shows a hydropathy plot for the predicted sequence of the mouse IGPcR17 protein compared to the human IGPcR17 protein and that of human KIAA0001. The analysis was performed using the method of Kyte and DooLittle (1982, J. Mol.

Biol., 157:105-32), with the DAMBE program (Data Analysis in Molecular Biology and Evolution), University of Hong Kong, version 3.7.49.

**Example 10. Generation of ES cells with a modified IGPcR17 allele, produced by homologous recombination.**

The most preferred method in this invention is described in Wattler S & Nehls M, German patent application DE 100 16 523.0, "Klonierungssystem zur Konstruktion von homologen Rekombiationsvektoren", filed April 03, 2000. This method reduces the time required for the construction of such a vector from 3-6 months to about 14 days. The vector includes a linear lambda vector (lambda-KO-Sfi) that comprises a stuffer fragment; an *E. coli* origin of replication; an antibiotic resistance gene for bacteria selection, two negative selection markers suitable for use in mammalian cells; LoxP sequences for cre-recombinase mediated conversion of linear lambda phages into high copy plasmids. In a final targeting vector, the stuffer fragment is replaced by nucleotide sequences representing a left arm of homology, an ES cell selection cassette, and a right arm of homology.

To abolish the gene function of mouse IGPcR17 (mIGPcR17) a deletion of approximately 880 bp of the coding region starting approximately 10 bp upstream of the ATG was performed (see Fig. 10). The left arm of homology (hereafter referred to as A/C) is PCR amplified with the primers C and A. The primers contain *Sfi I* restriction sites A and C in their 5'-ends, respectively. *Sfi* recognizes and cuts the nucleotide sequence 5-GGCCNNNNNGGCC-3'. By changing the nucleotides designated N, unique and non-compatible *Sfi* restriction sites are generated. The 3'-end of primer A is homologous to 25 bp of mouse IGPcR17, ending with the 10 bp downstream of the ATG. The 3'-end (25 bp) of primer C is homologous to a position

approximately 2500 basepairs upstream of the ATG. The right arm of homology (hereafter referred to as B/D) is PCR amplified with primers B and D: B is located approximately 880 bp downstream of the ATG, and D approximately 2000 bp downstream of the stop codon. Both primers contain *Sfi*-restriction sites B or D in their 5'-ends, respectively. To avoid the introduction of point mutations the Expand high fidelity PCR-System, (Boehringer Mannheim / Roche Diagnostics, Basel CH) is used. A ligation of A/C with B/D and a selection cassette leads to an approximately 880 bp deletion of the mIGPcR17 coding region, thereby creating a null allelle. Both PCR-products A/C and B/D are purified using Qiaquick PCR Purification Kit according to the manufacturer (Quiagen, Venlo, NL). The PCR-products are cleaved 3 hours at 50°C with 60 U *Sfi* and subsequently purified (Qiaquick PCR Purification kit). The final volume is 30 µl/product. The ES-cell selection cassette (IRES-ß-lactamase-MCSneo) contains *Sfi*-sites A and B 5'- and 3'-, respectively (Wattler S, *et al.*, 1999, Biotechniques, 26:1150-1159). A typical ligation is 50 ng lambda-KO-Sfi-arm (*Sfi*-cleaved), 10 ng selection cassette, 1 ng A/C, 1 ng B/D, 1 x ligation buffer and 1U T4 ligase (Boehringer Mannheim / Roche Diagnostics, Basel CH). The ligation is carried out for 2 hours at room temperature. Two µl of the ligation are used for *in vitro* packaging ('Gigapack plus' from Stratagene, La Jolla CA, USA) for 1.5 hours at room temperature according to the manufacturer's instructions. Aliquots of 10 µl and 50 µl are used to infect C600 bacteria (Stratagene, La Jolla CA, USA) and infection is performed overnight. Single plaques in SM-buffer (Ausubel FM *et al.*, 1994, "*Current Protocols in Molecular Biology*", John Wiley & Sons, New York) are taken to infect BNN 132 bacteria (30 min at 30°C) for plasmid conversion and infection. Bacteria are cultured over 16 hours at 30°C in TB media (Ausubel FM *et al.*, 1994, "*Current Protocols in Molecular Biology*", John Wiley & Sons, New York), containing 100 µg/ml ampicillin (Amersham Pharmacia Biotech, Piscataway NJ, USA; cat. no. US11259-25). Plasmids are harvested using the Qiagen plasmid kit (Qiagen cat. no. 12143) according to the manufacturer's instructions. To verify plasmid integrity, *Sfi* and *EcoR1*-digests are performed.

The transformation of mouse 129 ES cells with the final targeting vector is

performed according to standard procedures. Electroporated 129 mouse ES cells are double-selected with G418 (400 μg/ml) for 7 days and GANC (ganciclovir, 0.2 μM) for 3 days, starting on day 3 after electroporation, for positive and negative selection, respectively, thereby enriching for transformants having the neomycin resistance gene integrated into an endogenous IGPcR17 allele. Single cell clones are propagated, frozen down and expanded for DNA isolation. To identify positively targeted clones, ES cell DNA is isolated from selected clones, incubated with an appropriate restriction enzyme, and the digestion products separated on an agarose gel. Southern blots are hybridized with a 5' external probe and positive targeted candidates are verified by hybridization with a 3' external probe. A single integration is confirmed by hybridization with a probe derived from the neomycin gene. Positive ES cells are isolated and expanded in culture.

**Example 11. Mice Deficient in the Expression of the IGPcR17 Gene.**

Male chimeric mice are generated by micro-injection of ES cells carrying a recombined allele into 129/SvEv mouse blastocysts, using standard methodology. The chimeric blastocyst is implanted into the uterus of a pseudopregnant mouse, prepared by mating females with vasectomized males of the same species. The chimeras are bred to wild type animals. Tail DNA is isolated from the offspring of these chimeric mice and analyzed by incubation with appropriate restriction enzymes followed by Southern analysis, using the same strategy as outlined above to determine germline transmission. The blots demonstrate the transmission into the mouse genome of the mutation altering the IGPcR17 allele in transformant ES cells. The chimeric male mouse and its heterozygous progeny (+/-) are bred to produce mice homozygous for the mutation (-/-).

Northern blots are used to probe mRNA obtained from tissues of offspring for the presence or absence of transcripts encoding either the IGPcR27, the marker gene, or both. In addition, Western blots are used to assess IGPcR27 expression by probing

with antibody specific for the receptor.

Those skilled in the art will be able to recognize, or be able to ascertain, using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the invention.

## CLAIMS

1. An isolated nucleic acid molecule, wherein said nucleic acid molecule comprises at least one of:
   (a) the nucleotide sequence of SEQ ID NO:1;
   (b) the nucleotide sequence of SEQ ID NO:6;
   (c) a nucleotide sequence which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or any unique fragment thereof wherein the amino acid sequence of the fragment is greater than ten amino acids in length.
   (d) a nucleotide sequence which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:7, or any unique fragment thereof wherein the amino acid sequence of the fragment is greater than ten amino acids in length.

2. An isolated nucleic acid molecule comprising an allelic variant of a nucleotide sequence which encodes a polypeptide comprising the amino acid sequence selected from the group of:
   (a) the amino acid sequence of SEQ ID NO:2; and
   (b) the amino acid sequence of SEQ ID NO:7;
   wherein said allelic variant contains at least 80% nucleic acid homology and hybridizes to the complement of SEQ ID NO:1 under highly stringent conditions equivalent to hybridization in 42°C in a hybridization solution comprising 50% formamide, 1% SDS, 1M NaCl, 10% Dextran sulfate, and washing twice for 30 minutes in a wash solution comprising 0.1xSSC and 1% SDS.

3. The isolated nucleic acid molecule of claims 1 or 2, comprising a nucleotide sequence which encodes at least one of the group of polypeptides, peptides and fusion proteins, comprising an amino acid sequence at least 70% similar to an amino acid sequence selected from the group of:
   (a) the amino acid sequence of SEQ ID NO:2; and

(b) the amino acid sequence of SEQ ID NO:7.

4. The isolated nucleic acid molecule of claims 1 to 3 operatively linked with a nucleotide regulatory sequence capable of controlling expression of the nucleic acid molecule in a host cell or non-human animal.

5. A vector comprising the isolated nucleic acid molecule of any of claims 1 to 4.

6. A host cell genetically engineered to contain at least one of:

(a) the nucleic acid molecule of any of claims 1 to 4; or

(b) the vector of claim 5.

7. The host cell of claim 6 wherein said host cell is a eucaryotic cell, being at least one of:

(a) a yeast cell;

(b) an insect cell; or

(c) a mammalian cell.

8. The human IGPcR17 protein of SEQ ID NO:2, or any unique fragment thereof wherein the amino acid sequence of the fragment is greater than ten amino acids in length, including but not limited to polypeptides, peptides, isolated domains and fusion proteins.

9. The mouse IGPcR17 protein of SEQ ID NO:7, or any unique fragment thereof wherein the amino acid sequence of the fragment is greater than ten amino acids in length, including but not limited to polypeptides, peptides, isolated domains and fusion proteins.

10. Antibodies specifically targeting the IGPcR17 proteins of any of claims 8 or 9, and/or polypeptides, peptides, isolated domains and the the IGPcR17 component of fusion proteins of said IGPcR17 proteins.

11. Agonists and antagonists of IGPcR17 protein that compete selectively with native natural IGPcR17 ligand and which modulate IGPcR17 gene expression or gene product activity, including: (a) 'small molecules' of molecular mass less than 6 kDa; (b) molecules of intermediate size, having molecular mass between 5 kDa to 15 kDa; and (c) large molecules of molecular mass greater than 12 kDa; the latter including mutant natural IGPcR17 ligand proteins that compete with native natural IGPcR17 ligand and which modulate IGPcR17 gene expression or gene product activity.

12. Anti-sense and ribozyme molecules that can be used to inhibit IGPcR17 gene expression or expression constructs used to enhance IGPcR17 gene expression.

13. Methods of identifying compounds of any of claims 11 or 12, which modulate the activity of IGPcR17 or IGPcR17 gene expression.

14. Embryonic stem cells containing a disrupted endogenous IGPcR17 gene.

15. Non-human knock-out animals that do not express IGPcR17, wherein the endogenous animal ortholog of the IGPcR17 gene is functionally disrupted.

16. The non-human knock-out animals of claim 15, wherein the endogenous animal ortholog of the IGPcR17 gene is functionally disrupted by an homologous recombination method.

17. Mutated non-human animals that express a non-functional or partially functional form of IGPcR17.

18. A non-human transgenic animal model expressing the human IGPcR17 cDNA sequence as shown in SEQ ID NO:1 or the nucleic acid molecule of any of claims 1 to 4.

19. The non-human animal model according to any one of claims 17 to 18, whereby the human IGPcR17 is encoded by a nucleic acid sequence which is homozygous in said animal model.

20. Progeny of non-human animals of any of claims 15 to 19, including both heterozygous and homozygous offspring.

21. Non-human animals of any of claims 15 to 20, wherein the animal is from a genus selected from the group consisting of *Mus* (*e.g.*, mice), *Rattus* (*e.g.*, rats), *Oryctologus* (*e.g.*, rabbits) and *Mesocricetus* (*e.g.*, hamsters).

22. Use of the non-human animal according to any one of claims 15 to 21, for the dissection of the molecular mechanisms of the IGPcR17 pathway, for the identification and cloning of genes able to modify, reduce or inhibit the phenotype associated with IGPcR17 activity or deficiency.

23. Use of the animal model according to any of claims 15 to 21 for the identification of gene and protein diagnostic markers for diseases.

24. Use of the animal model according to any of claims 15 to 21 for the identification and testing of compounds useful in the prevention, amelioration or treatment of diseases associated with IGPcR17 activity or deficiency.

25. The use of any of claims 23 or 24 wherein the disease is selected from the group of psychiatric and central nervous system diseases associated with signal processing in the central nervous system.

26. The use of claims 25 wherein the disease is selected from the group of learning and memory disorders, movement dysfunctions, tics, tremor, Tourette's syndrome, Parkinson's disease, Huntington's disease, dyskinesias, dystonia, pain and spasms.

27. A method of identifying compounds suitable for modulating the activity of the protein according to claim 8, for treatment of diseases characterized by aberrant expression or activity of IGPcR17.

28. A method of prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17, by the administration of compounds that bind specifically to the IGPcR17 gene or protein and/or which modulate IGPcR17 expression or IGPcR17 activity; the compounds that that bind specifically to the IGPcR17 gene or protein and/or which modulate IGPcR17 expression or IGPcR17 activity for the prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17; and the use of compounds that that bind specifically to the IGPcR17 gene or protein and/or which modulate IGPcR17 expression or IGPcR17 activity for prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17.

29. A gene therapy method of prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17, by the administration of vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7, that modulate IGPcR17 expression or IGPcR17 activity; the vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7 which modulate IGPcR17 expression or IGPcR17 activity for the prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17; and the use of vectors and/or host cells containing nucleotide sequences according to any of claims 1 to 7 which modulate IGPcR17 expression or IGPcR17 activity for prevention, amelioration or treatment of diseases characterized by aberrant expression or activity of IGPcR17.

30. The method of any of claims 27 to 29 wherein the disease is selected from the group of psychiatric and central nervous system diseases associated with signal processing in the central nervous system.

31. The method of claim 30 wherein the disease is selected from the group of learning and memory disorders, movement dysfunctions, tics, tremor, Tourette's syndrome, Parkinson's disease, Huntington's disease, dyskinesias, dystonia, pain and spasms.

Figure 1. **Human IGPcR17 cDNA sequence.**

ATGCAAGCCGTCGACAACCTCACCTCTGCGCCTGGGAACACCAGTCTGTGCACCAGAGACTACAAA
ATCACCCAGGTCCTCTTCCCACTGCTCTACACTGTCCTGTTTTTTGTTGGACTTATCACAAATGGCCT
GGCGATGAGGATTTTCTTTCAAATCCGGAGTAAATCAAACTTTATTATTTTTCTTAAGAACACAGTC
ATTTCTGATCTTCTCATGATTCTGACTTTTCCATTCAAAATTCTTAGTGATGCCAAACTGGGAACAG
GACCACTGAGAACTTTTGTGTGTCAAGTTACCTCCGTCATATTTTATTTCACAATGTATATCAGTAT
TTCATTCCTGGGACTGATAACTATCGATCGCTACCAGAAGACCACCAGGCCATTTAAAACATCCAA
CCCCAAAAATCTCTTGGGGGCTAAGATTCTCTCTGTTGTCATCTGGGCATTCATGTTCTTACTCTCTT
TGCCTAACATGATTCTGACCAACAGGCAGCCGAGAGACAAGAATGTGAAGAAATGCTCTTTCCTTA
AATCAGAGTTCGGTCTAGTCTGGCATGAAATAGTAAATTACATCTGTCAAGTCATTTTCTGGATTAA
TTTCTTAATTGTTATTGTATGTTATACACTCATTACAAAAGAACTGTACCGGTCATACGTAAGAACG
AGGGGTGTAGGTAAAGTCCCCAGGAAAAAGGTGAACGTCAAAGTTTTCATTATCATTGCTGTATTC
TTTATTTGTTTTGTTCCTTTCCATTTTGCCCGAATTCCTTACACCCTGAGCCAAACCCGGGATGTCTT
TGACTGCACTGCTGAAAATACTCTGTTCTATGTGAAAGAGAGCACTCTGTGGTTAACTTCCTTAAAT
GCATGCCTGGATCCGTTCATCTATTTTTTCCTTTGCAAGTCCTTCAGAAATTCCTTGATAAGTATGCT
GAAGTGCCCCAATTCTGCAACATCTCTGTCCCAGGACAATAGGAAAAAAGAACAGGATGGTGGTG
ACCCAAATGAAGAGACTCCAATG

Figure 2. **Human IGPcR17 amino acid sequence.**

MQAVDNLTSAPGNTSLCTRDYKITQVLFPLLYTVLFFVGLITNGLAMRIFFQIRSKSNFIIFLKNTVISDLL
MILTFPFKILSDAKLGTGPLRTFVCQVTSVIFYFTMYISISFLGLITIDRYQKTTRPFKTSNPKNLLGAKILS
VVIWAFMFLLSLPNMILTNRQPRDKNVKKCSFLKSEFGLVWHEIVNYICQVIFWINFLIVIVCYTLITKEL
YRSYVRTRGVGKVPRKKVNVKVFIIIAVFFICFVPFHFARIPYTLSQTRDVFDCTAENTLFYVKESTLWLT
SLNACLDPFIYFFLCKSFRNSLISMLKCPNSATSLSQDNRKKEQDGGDPNEETPM

Figure 3A. Northern Blot Analysis of human IGPcR17

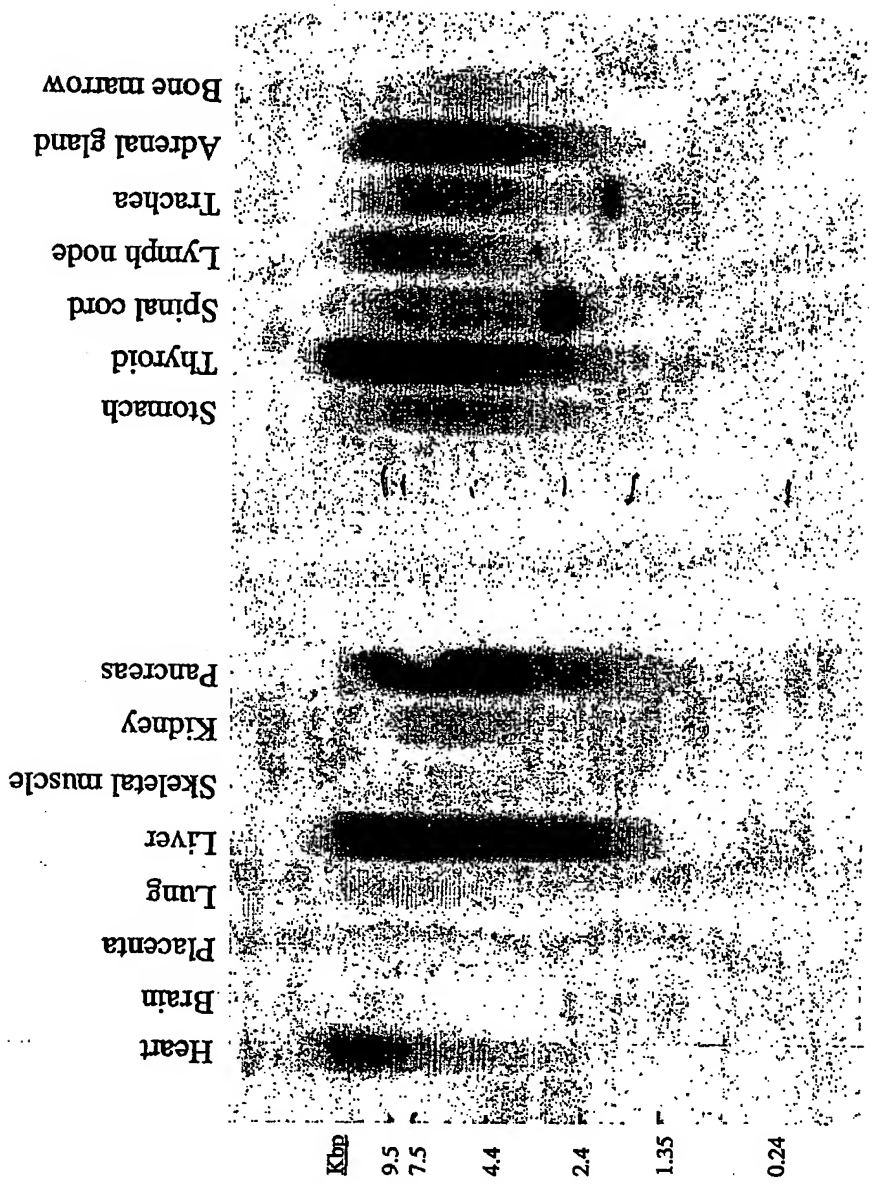


Figure 3B. Northern Blot Analysis of human IGPcR17.

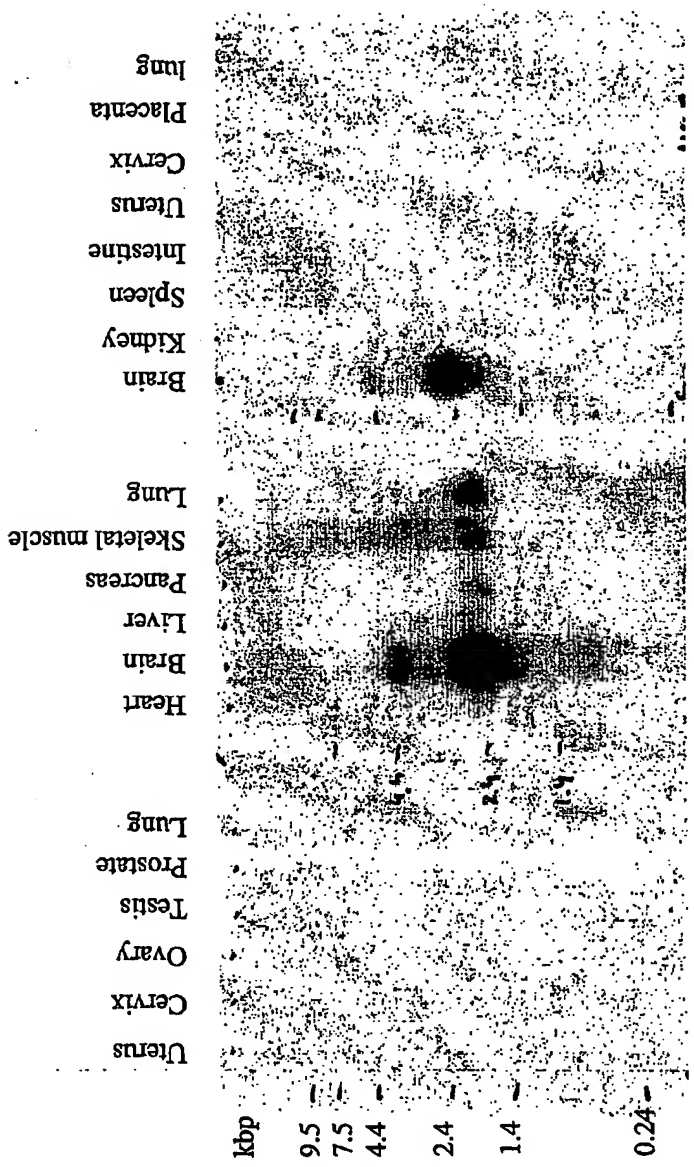


Figure 4a. Dot Blot Analysis of human IGPcR17.

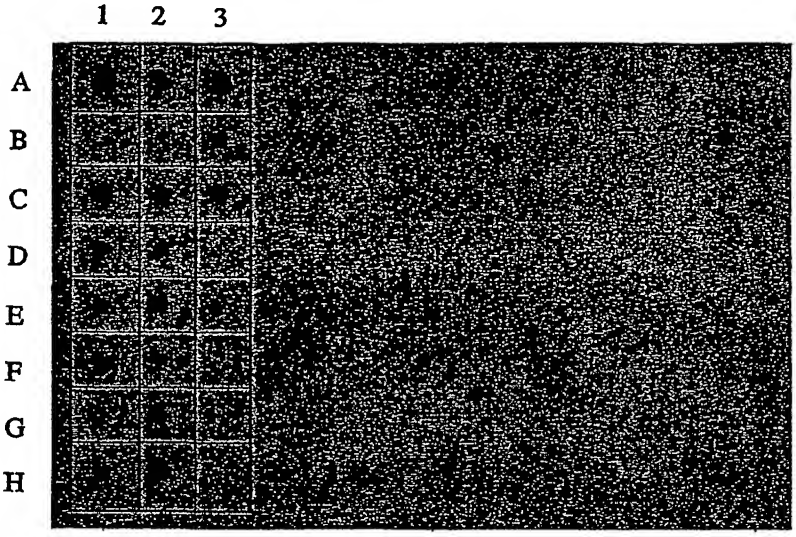



Figure 4b.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | whole brain | cerebellum, left | substantia nigra | heart | esophagus | colon, transverse | kidney | lung | liver | leukemia, HL-60 | fetal brain | yeast total RNA |
| B | cerebral cortex | cerebellum, right | accumbens nucleus | aorta | stomach | colon, descending | skeletal muscle | placenta | pancreas | HeLa S3 | fetal heart | yeast tRNA |
| C | frontal lobe | corpus callosum | thalamus | atrium, left | duodenum | rectum | spleen | bladder | adrenal gland | leukemia, K-562 | fetal kidney | *E. coli* rRNA |
| D | parietal lobe | amygdala | pituitary gland | atrium, right | jejunum | | thymus | uterus | thyroid gland | leukemia, RPMI-8226 | fetal liver | *E. coli* DNA |
| E | occipital lobe | caudate, nucleus | spinal cord | ventricle, left | ileum | | peripheral blood leukocyte | prostate | salivary gland | Burkitt's lymphoma, Raji | fetal spleen | Poly r(A) |
| F | temporal lobe | hippo-campus | | ventricle, right | ileocecum | | lymph node | testis | mammary gland | Burkitt's lymphoma, Daudi | fetal thymus | human $C_0t$-1 DNA |
| G | p. g.* of cerebral cortex | medulla oblongata | | inter-ventricular septum | appendix | | bone marrow | ovary | | colorectal adeno-carcinoma, SW480 | fetal lung | human DNA 100 ng |
| H | pons | putamen | | apex of the heart | colon, ascending | | trachea | | | lung carcinoma, A549 | | human DNA 500 ng |

* paracentral gyrus

# Figure 5.

**Mouse IGPcR17 cDNA sequence**

ATGGATGTGCCTGGTGTCAACACCACCTCAGCCAATACCACCTTCTCCCCTGGGACCAGCACCCTG
TGCGTCAGAGACTACAAGATCACCCAGGTTCTCTTCCCATTGCTGTACACCGTCCTGTTCTTTGCTG
GGCTCATCACGAACAGCTTGGCAATGAGGATTTTCTTTCAGATCCGCAGTAAATCCAACTTCATCAT
TTTTCTTAAGAACACGGTCATCTCTGATCTACTAATGATTCTAACTTTTCCATTTAAAATTCTTAGTG
ATGCTAAACTGGGAGCCGGGCCTCTGAGAACCTTGGTGTGCCAAGTTACTTCAGTCACATTTTATTT
TACAATGTATATCAGTATATCGTTCCTGGGGTTGATAACCATTGACCGCTACCTGAAGACCACCAG
GCCATTTAAAACGTCCAGCCCCAGCAATCTCTTGGGTGCAAAGATTCTTTCTGTTGTCATCTGGGCC
TTCATGTTCTTAATTTCACTGCCTAACATGATTCTCACCAACAGGAGGCCAAAAGATAAGGACGTA
ACAAAATGTTCTTTCTTAAAGTCAGAGTTTGGTCTAGTTTGGCACGAAATAGTCAATTACATCTGCC
AAGTCATTTTCTGGATTAATTTTTTAATTGTCATCGTTTGTTATAGCCTCATTACCAAAGAACTCTAT
CGGTCTTATGTCAGAACAAGGGGTTCAGCCAAAGTTCCCAAGAAAAAGGTAAACGTCAAGGTTTTC
ATCATCATTGCTGTATTCTTTATTTGCTTTGTTCCCTTCCACTTTGCACGGATTCCCTACACCCTGAG
CCAAACTCGGGCCGTCTTTGACTGCAGTGCTGAGAACACCCTGTTCTACGTGAAGGAGAGCACCCT
ATGGCTGACGTCACTGAACGCCTGCCTTGATCCATTCATCTACTTTTTTCTTTGCAAGTCTTTCAGAA
ATTCCTTGACAAGCATGCTGAGGTGCTCAAACTCTACATCAACATCTGGGACAAACAAGAAGAAA
GGACAAGAAGGTGGCGAACCAAGCGAAGAGACCCCAATGTAG

# Figure 6.

**Mouse IGPcR17 amino acid sequence**

MDVPGVNTTSANTTFSPGTSTLCVRDYKITQVLFPLLYTVLFFAGLITNSLAMRIFFQIRSKSNFIIFLKNT
VISDLLMILTFPFKILSDAKLGAGPLRTLVCQVTSVTFYFTMYISISFLGLITIDRYLKTTRPFKTSSPSNLL
GAKILSVVIWAFMFLISLPNMILTNRRPKDKDVTKCSFLKSEFGLVWHEIVNYICQVIFWINFLIVIVCYSL
ITKELYRSYVRTRGSAKVPKKKVNVKVFIIIAVFFICFVPFHFARIPYTLSQTRAVFDCSAENTLFYVKEST
LWLTSLNACLDPFIYFFLCKSFRNSLTSMLRCSNSTSTSGTNKKKGQEGGEPSEETPM

Figure 7a. Amino acid alignment of human IGPcR17 and mouse IGPcR17.

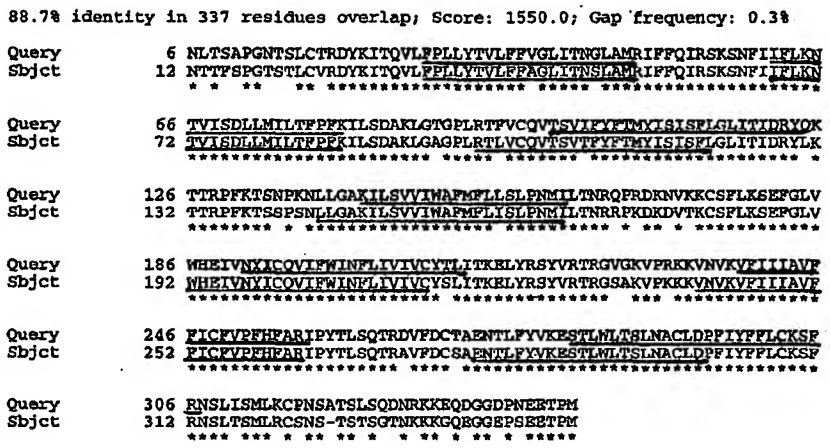
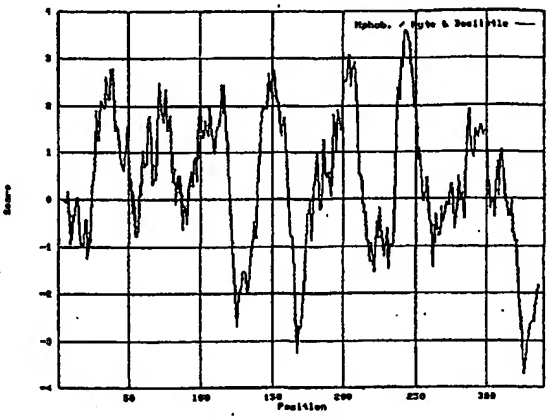


Figure 7b. Amino acid alignment of human IGPcR17 and human KIAA0001.

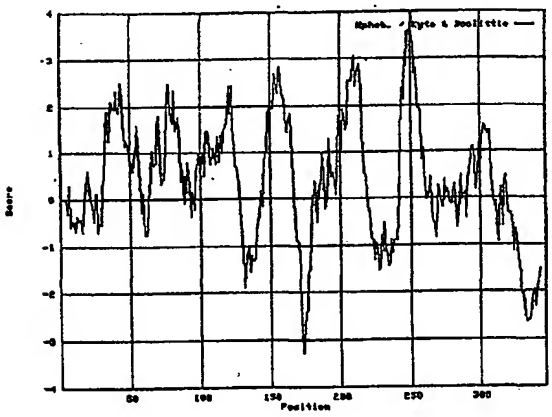
```
Score =  305 bits (772), Expect = 1e-87
 Identities = 149/314 (47%), Positives = 209/314 (66%)

Query: 6   NLTSAPGNTSLCTRDYKITQVLFPLLYTVLFFVGLITNGLAMRIFFQIRSKSNFIIFLKN 65
           N TS       C+++  ITQ + P+LY ++F  G++ NG++  IFF + S  +FII+LKN
Sbjct: 3   NSTSTQPPDESCSQNLLITQQIIPVLYCMVFIAGILLNGVSGWIFFYVPSSKSFIIYLKN 62

Query: 66  TVISDLLMILTFPFKILSDAKLGTGPLRTFVCQVTSVIFYFTMYISISFLGLITIDRYQK 125
            VI+D +M LTFPFKIL D+ LG   L  FVC+V++V+FY  MY+SI F GLI+ DRY K
Sbjct: 63  IVIADFVMSLTFPFKILGDSGLGPWQLNVFVCRVSAVLFYVNMYVSIVFFGLISFDRYYK 122

Query: 126 TTRPFKTSNPKNLLGAKILSVVIWAFMFLLSLPNMILTNRQPRDKNVKKCSFLKSEFGLV 185
             +P  TS  +++  +K+LSV++W  M LL++PN+ILTN+  R+   . KC  LKSE G
Sbjct: 123 IVKPLWTSFIQSVSYSKLLSVIVWMLMLLLAVPNIILTNQSVREVTQIKCIELKSELGRK 182

Query: 186 WHEIVNYICQVIFWINFLIVIVCYTLITKELYRSYVRTRGVGKVPRKKVNVKVFIIIAVF 245
           WH+  NYI   IFWI FL++IV YT ITK++++S++++        +KK +  +F I+ VF
Sbjct: 183 WHKASNYIFVAIFWIVFLLLIVFYTAITKKIFKSHLKSSRNSTSVKKKSSRNIFSIVFVF 242

Query: 246 FICFVPFHFARIPYTLSQTRDVFDCTAENTLFYVKESTLWLTSLNACLDPFIYFFLCKSF 305
           F+CFVP+H ARIPYT SQT   + C ++  L Y+KE TL L++ N CLDP IYFFLC+ F
Sbjct: 243 FVCFVPYHIARIPYTKSQTEAHYSCQSKEILRYMKEFTLLLSAANVCLDPIIYFFLCQPF 302

Query: 306 RNSLISMLKCPNSA 319
           R L   L P A
Sbjct: 303 REILCKKLHIPLKA 316
```

Figure 8. Hydropathy plots, comparing human IGPcR17, mouse IGPcR17 and human KIAA0001.

human IGPcR17



mouse IGPcR17



human KIAA0001

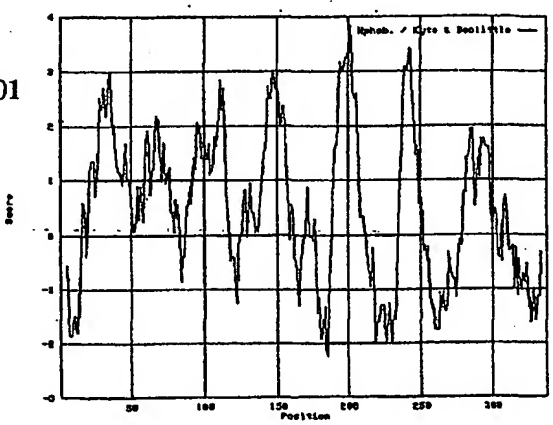


Figure 9. Northern Blot Analysis of mouse IGPcR17.

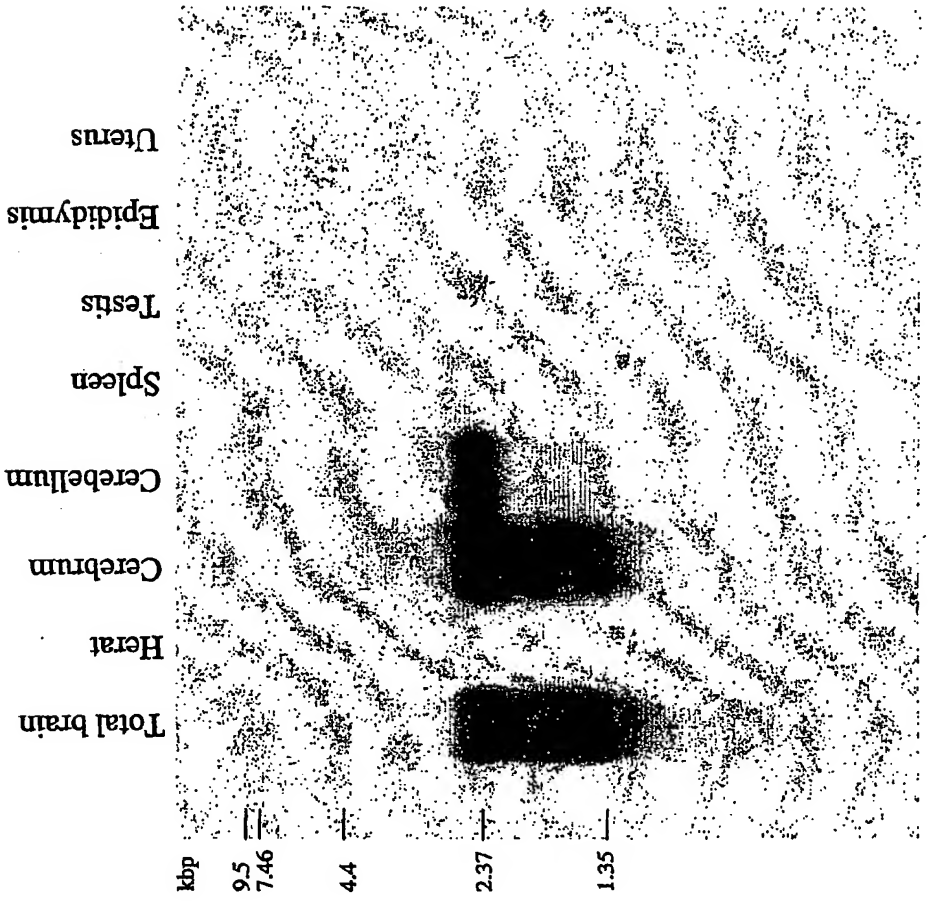
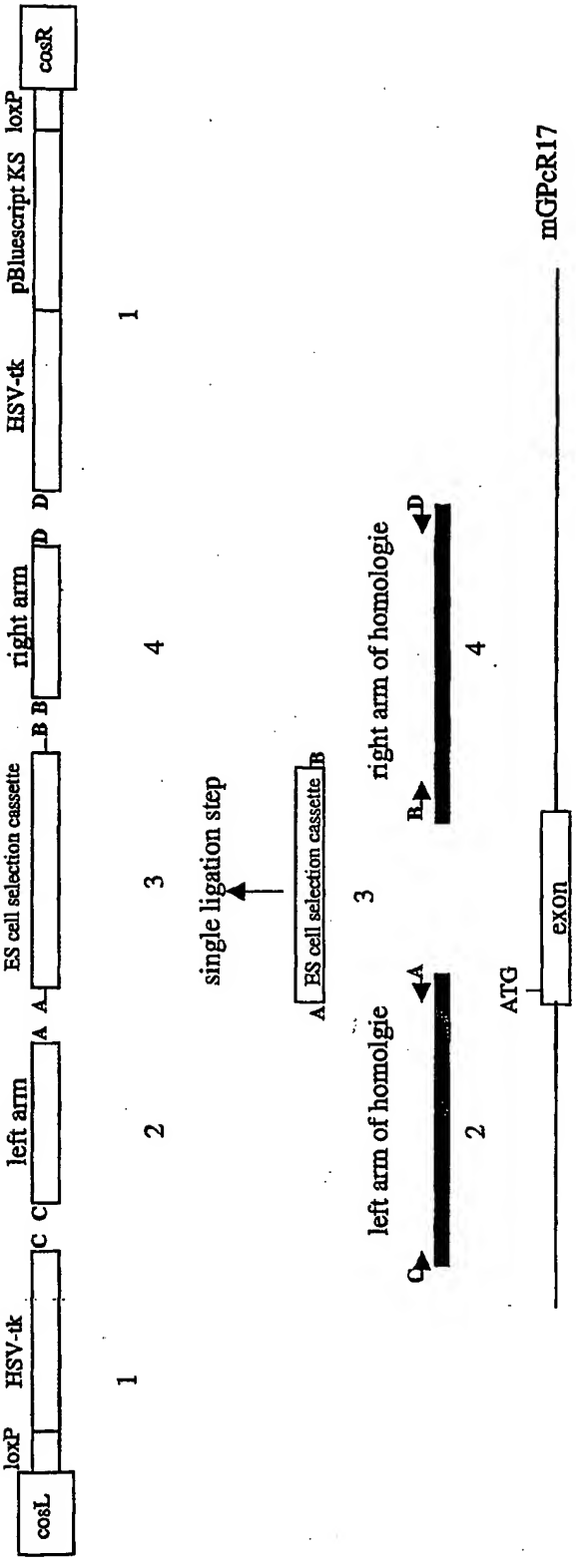


Figure 10. Targeting vector construction.



SEQUENCE LISTING

<110> Ingenium Pharmaceuticals AG

<120> Human G protein-coupled receptor, IGPcR17, and uses thereof

<130> IPA-004-EWO

<140>
<141>

<150> US 60/215,759
<151> 2000-06-30

<160> 10

<170> PatentIn Ver. 2.1

<210> 1
<211> 1029
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1026)

<300>
<303> Nature
<304> 409
<306> 202-207
<307> 2001-01-11
<308> EMBL/313449
<309> 2001-01-12

<400> 1

```
atg caa gcc gtc gac aac ctc acc tct gcg cct ggg aac acc agt ctg   48
Met Gln Ala Val Asp Asn Leu Thr Ser Ala Pro Gly Asn Thr Ser Leu
 1               5                  10                  15

tgc acc aga gac tac aaa atc acc cag gtc ctc ttc cca ctg ctc tac   96
Cys Thr Arg Asp Tyr Lys Ile Thr Gln Val Leu Phe Pro Leu Leu Tyr
            20                  25                  30

act gtc ctg ttt ttt gtt gga ctt atc aca aat ggc ctg gcg atg agg   144
Thr Val Leu Phe Phe Val Gly Leu Ile Thr Asn Gly Leu Ala Met Arg
        35                  40                  45
```

```
att ttc ttt caa atc cgg agt aaa tca aac ttt att att ttt ctt aag   192
Ile Phe Phe Gln Ile Arg Ser Lys Ser Asn Phe Ile Ile Phe Leu Lys
     50                  55                  60

aac aca gtc att tct gat ctt ctc atg att ctg act ttt cca ttc aaa   240
Asn Thr Val Ile Ser Asp Leu Leu Met Ile Leu Thr Phe Pro Phe Lys
 65                  70                  75                  80

att ctt agt gat gcc aaa ctg gga aca gga cca ctg aga act ttt gtg   288
Ile Leu Ser Asp Ala Lys Leu Gly Thr Gly Pro Leu Arg Thr Phe Val
                 85                  90                  95

tgt caa gtt acc tcc gtc ata ttt tat ttc aca atg tat atc agt att   336
Cys Gln Val Thr Ser Val Ile Phe Tyr Phe Thr Met Tyr Ile Ser Ile
            100                 105                 110

tca ttc ctg gga ctg ata act atc gat cgc tac cag aag acc acc agg   384
Ser Phe Leu Gly Leu Ile Thr Ile Asp Arg Tyr Gln Lys Thr Thr Arg
        115                 120                 125

cca ttt aaa aca tcc aac ccc aaa aat ctc ttg ggg gct aag att ctc   432
Pro Phe Lys Thr Ser Asn Pro Lys Asn Leu Leu Gly Ala Lys Ile Leu
    130                 135                 140

tct gtt gtc atc tgg gca ttc atg ttc tta ctc tct ttg cct aac atg   480
Ser Val Val Ile Trp Ala Phe Met Phe Leu Leu Ser Leu Pro Asn Met
145                 150                 155                 160

att ctg acc aac agg cag ccg aga gac aag aat gtg aag aaa tgc tct   528
Ile Leu Thr Asn Arg Gln Pro Arg Asp Lys Asn Val Lys Lys Cys Ser
                165                 170                 175

ttc ctt aaa tca gag ttc ggt cta gtc tgg cat gaa ata gta aat tac   576
Phe Leu Lys Ser Glu Phe Gly Leu Val Trp His Glu Ile Val Asn Tyr
            180                 185                 190

atc tgt caa gtc att ttc tgg att aat ttc tta att gtt att gta tgt   624
Ile Cys Gln Val Ile Phe Trp Ile Asn Phe Leu Ile Val Ile Val Cys
        195                 200                 205

tat aca ctc att aca aaa gaa ctg tac cgg tca tac gta aga acg agg   672
Tyr Thr Leu Ile Thr Lys Glu Leu Tyr Arg Ser Tyr Val Arg Thr Arg
    210                 215                 220

ggt gta ggt aaa gtc ccc agg aaa aag gtg aac gtc aaa gtt ttc att   720
Gly Val Gly Lys Val Pro Arg Lys Lys Val Asn Val Lys Val Phe Ile
225                 230                 235                 240
```

```
atc att gct gta ttc ttt att tgt ttt gtt cct ttc cat ttt gcc cga   768
Ile Ile Ala Val Phe Phe Ile Cys Phe Val Pro Phe His Phe Ala Arg
                245                 250                 255

att cct tac acc ctg agc caa acc cgg gat gtc ttt gac tgc act gct   816
Ile Pro Tyr Thr Leu Ser Gln Thr Arg Asp Val Phe Asp Cys Thr Ala
            260                 265                 270

gaa aat act ctg ttc tat gtg aaa gag agc act ctg tgg tta act tcc   864
Glu Asn Thr Leu Phe Tyr Val Lys Glu Ser Thr Leu Trp Leu Thr Ser
        275                 280                 285

tta aat gca tgc ctg gat ccg ttc atc tat ttt ttc ctt tgc aag tcc   912
Leu Asn Ala Cys Leu Asp Pro Phe Ile Tyr Phe Phe Leu Cys Lys Ser
    290                 295                 300

ttc aga aat tcc ttg ata agt atg ctg aag tgc ccc aat tct gca aca   960
Phe Arg Asn Ser Leu Ile Ser Met Leu Lys Cys Pro Asn Ser Ala Thr
305                 310                 315                 320

tct ctg tcc cag gac aat agg aaa aaa gaa cag gat ggt ggt gac cca   1008
Ser Leu Ser Gln Asp Asn Arg Lys Lys Glu Gln Asp Gly Gly Asp Pro
                325                 330                 335

aat gaa gag act cca atg taa                                     1029
Asn Glu Glu Thr Pro Met
            340
```

<210> 2
<211> 342
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Gln Ala Val Asp Asn Leu Thr Ser Ala Pro Gly Asn Thr Ser Leu
  1               5                  10                  15

Cys Thr Arg Asp Tyr Lys Ile Thr Gln Val Leu Phe Pro Leu Leu Tyr
             20                  25                  30

Thr Val Leu Phe Phe Val Gly Leu Ile Thr Asn Gly Leu Ala Met Arg
         35                  40                  45

Ile Phe Phe Gln Ile Arg Ser Lys Ser Asn Phe Ile Ile Phe Leu Lys
     50                  55                  60
```

```
Asn Thr Val Ile Ser Asp Leu Leu Met Ile Leu Thr Phe Pro Phe Lys
 65                  70                  75                  80

Ile Leu Ser Asp Ala Lys Leu Gly Thr Gly Pro Leu Arg Thr Phe Val
                85                  90                  95

Cys Gln Val Thr Ser Val Ile Phe Tyr Phe Thr Met Tyr Ile Ser Ile
            100                 105                 110

Ser Phe Leu Gly Leu Ile Thr Ile Asp Arg Tyr Gln Lys Thr Thr Arg
        115                 120                 125

Pro Phe Lys Thr Ser Asn Pro Lys Asn Leu Leu Gly Ala Lys Ile Leu
    130                 135                 140

Ser Val Val Ile Trp Ala Phe Met Phe Leu Leu Ser Leu Pro Asn Met
145                 150                 155                 160

Ile Leu Thr Asn Arg Gln Pro Arg Asp Lys Asn Val Lys Lys Cys Ser
                165                 170                 175

Phe Leu Lys Ser Glu Phe Gly Leu Val Trp His Glu Ile Val Asn Tyr
            180                 185                 190

Ile Cys Gln Val Ile Phe Trp Ile Asn Phe Leu Ile Val Ile Val Cys
        195                 200                 205

Tyr Thr Leu Ile Thr Lys Glu Leu Tyr Arg Ser Tyr Val Arg Thr Arg
    210                 215                 220

Gly Val Gly Lys Val Pro Arg Lys Lys Val Asn Val Lys Val Phe Ile
225                 230                 235                 240

Ile Ile Ala Val Phe Phe Ile Cys Phe Val Pro Phe His Phe Ala Arg
                245                 250                 255

Ile Pro Tyr Thr Leu Ser Gln Thr Arg Asp Val Phe Asp Cys Thr Ala
            260                 265                 270

Glu Asn Thr Leu Phe Tyr Val Lys Glu Ser Thr Leu Trp Leu Thr Ser
        275                 280                 285

Leu Asn Ala Cys Leu Asp Pro Phe Ile Tyr Phe Phe Leu Cys Lys Ser
    290                 295                 300

Phe Arg Asn Ser Leu Ile Ser Met Leu Lys Cys Pro Asn Ser Ala Thr
305                 310                 315                 320
```

Ser Leu Ser Gln Asp Asn Arg Lys Lys Glu Gln Asp Gly Gly Asp Pro
325 330 335

Asn Glu Glu Thr Pro Met
340

<210> 3
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: human oligonucleotide

<400> 3
cactgctcta cactgtcctg 20

<210> 4
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: human oligonucleotide

<400> 4
gaccgaactc tgatttaagg 20

<210> 5
<211> 466
<212> DNA
<213> Homo sapiens

<400> 5
ccactgctct acactgtcct gtttttgtt ggacttatca caaatggcct ggcgatgagg 60
attttctttc aaatccggag taaatcaaac tttattattt ttcttaagaa cacagtcatt 120
tctgatcttc tcatgattct gacttttcca ttcaaaattc ttagtgatgc caaactggga 180
acaggaccac tgagaacttt tgtgtgtcaa gttacctccg tcatatttta tttcacaatg 240
tatatcagta tttcattcct gggactgata actatcgatc gctaccagaa gaccaccagg 300
ccatttaaaa catccaaccc caaaaatctc ttgggggcta agattctctc tgttgtcatc 360
tgggcattca tgttcttact ctctttgcct aacatgattc tgaccaacag gcagccgaga 420
gacaagaatg tgaagaaatg ctctttcctt aaatcagagt tcggtc 466

<210> 6
<211> 1044
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (1)..(1041)

<300>
<302> mouse EST
<308> GeneBank/AA274112.1
<309> 1997-03-28
<313> 287 TO 829

<400> 6

```
atg gat gtg cct ggt gtc aac acc acc tca gcc aat acc acc ttc tcc   48
Met Asp Val Pro Gly Val Asn Thr Thr Ser Ala Asn Thr Thr Phe Ser
  1               5                  10                  15

cct ggg acc agc acc ctg tgc gtc aga gac tac aag atc acc cag gtt   96
Pro Gly Thr Ser Thr Leu Cys Val Arg Asp Tyr Lys Ile Thr Gln Val
             20                  25                  30

ctc ttc cca ttg ctg tac acc gtc ctg ttc ttt gct ggg ctc atc acg   144
Leu Phe Pro Leu Leu Tyr Thr Val Leu Phe Phe Ala Gly Leu Ile Thr
         35                  40                  45

aac agc ttg gca atg agg att ttc ttt cag atc cgc agt aaa tcc aac   192
Asn Ser Leu Ala Met Arg Ile Phe Phe Gln Ile Arg Ser Lys Ser Asn
     50                  55                  60

ttc atc att ttt ctt aag aac acg gtc atc tct gat cta cta atg att   240
Phe Ile Ile Phe Leu Lys Asn Thr Val Ile Ser Asp Leu Leu Met Ile
 65                  70                  75                  80

cta act ttt cca ttt aaa att ctt agt gat gct aaa ctg gga gcc ggg   288
Leu Thr Phe Pro Phe Lys Ile Leu Ser Asp Ala Lys Leu Gly Ala Gly
                 85                  90                  95

cct ctg aga acc ttg gtg tgc caa gtt act tca gtc aca ttt tat ttt   336
Pro Leu Arg Thr Leu Val Cys Gln Val Thr Ser Val Thr Phe Tyr Phe
            100                 105                 110

aca atg tat atc agt ata tcg ttc ctg ggg ttg ata acc att gac cgc   384
Thr Met Tyr Ile Ser Ile Ser Phe Leu Gly Leu Ile Thr Ile Asp Arg
```

```
        115                 120                 125

tac ctg aag acc acc agg cca ttt aaa acg tcc agc ccc agc aat ctc   432
Tyr Leu Lys Thr Thr Arg Pro Phe Lys Thr Ser Ser Pro Ser Asn Leu
    130                 135                 140

ttg ggt gca aag att ctt tct gtt gtc atc tgg gcc ttc atg ttc tta   480
Leu Gly Ala Lys Ile Leu Ser Val Val Ile Trp Ala Phe Met Phe Leu
145                 150                 155                 160

att tca ctg cct aac atg att ctc acc aac agg agg cca aaa gat aag   528
Ile Ser Leu Pro Asn Met Ile Leu Thr Asn Arg Arg Pro Lys Asp Lys
                165                 170                 175

gac gta aca aaa tgt tct ttc tta aag tca gag ttt ggt cta gtt tgg   576
Asp Val Thr Lys Cys Ser Phe Leu Lys Ser Glu Phe Gly Leu Val Trp
            180                 185                 190

cac gaa ata gtc aat tac atc tgc caa gtc att ttc tgg att aat ttt   624
His Glu Ile Val Asn Tyr Ile Cys Gln Val Ile Phe Trp Ile Asn Phe
        195                 200                 205

tta att gtc atc gtt tgt tat agc ctc att acc aaa gaa ctc tat cgg   672
Leu Ile Val Ile Val Cys Tyr Ser Leu Ile Thr Lys Glu Leu Tyr Arg
    210                 215                 220

tct tat gtc aga aca agg ggt tca gcc aaa gtt ccc aag aaa aag gta   720
Ser Tyr Val Arg Thr Arg Gly Ser Ala Lys Val Pro Lys Lys Lys Val
225                 230                 235                 240

aac gtc aag gtt ttc atc atc att gct gta ttc ttt att tgc ttt gtt   768
Asn Val Lys Val Phe Ile Ile Ile Ala Val Phe Phe Ile Cys Phe Val
                245                 250                 255

ccc ttc cac ttt gca cgg att ccc tac acc ctg agc caa act cgg gcc   816
Pro Phe His Phe Ala Arg Ile Pro Tyr Thr Leu Ser Gln Thr Arg Ala
            260                 265                 270

gtc ttt gac tgc agt gct gag aac acc ctg ttc tac gtg aag gag agc   864
Val Phe Asp Cys Ser Ala Glu Asn Thr Leu Phe Tyr Val Lys Glu Ser
        275                 280                 285

acc cta tgg ctg acg tca ctg aac gcc tgc ctt gat cca ttc atc tac   912
Thr Leu Trp Leu Thr Ser Leu Asn Ala Cys Leu Asp Pro Phe Ile Tyr
    290                 295                 300

ttt ttt ctt tgc aag tct ttc aga aat tcc ttg aca agc atg ctg agg   960
Phe Phe Leu Cys Lys Ser Phe Arg Asn Ser Leu Thr Ser Met Leu Arg
```

```
305                 310                 315                 320

tgc tca aac tct aca tca aca tct ggg aca aac aag aag aaa gga caa   1008
Cys Ser Asn Ser Thr Ser Thr Ser Gly Thr Asn Lys Lys Lys Gly Gln
                325                 330                 335

gaa ggt ggc gaa cca agc gaa gag acc cca atg tag                   1044
Glu Gly Gly Glu Pro Ser Glu Glu Thr Pro Met
            340                 345


<210> 7
<211> 347
<212> PRT
<213> Mus musculus

<400> 7
Met Asp Val Pro Gly Val Asn Thr Thr Ser Ala Asn Thr Thr Phe Ser
  1               5                  10                  15

Pro Gly Thr Ser Thr Leu Cys Val Arg Asp Tyr Lys Ile Thr Gln Val
             20                  25                  30

Leu Phe Pro Leu Leu Tyr Thr Val Leu Phe Phe Ala Gly Leu Ile Thr
         35                  40                  45

Asn Ser Leu Ala Met Arg Ile Phe Phe Gln Ile Arg Ser Lys Ser Asn
     50                  55                  60

Phe Ile Ile Phe Leu Lys Asn Thr Val Ile Ser Asp Leu Leu Met Ile
 65                  70                  75                  80

Leu Thr Phe Pro Phe Lys Ile Leu Ser Asp Ala Lys Leu Gly Ala Gly
                85                  90                  95

Pro Leu Arg Thr Leu Val Cys Gln Val Thr Ser Val Thr Phe Tyr Phe
            100                 105                 110

Thr Met Tyr Ile Ser Ile Ser Phe Leu Gly Leu Ile Thr Ile Asp Arg
        115                 120                 125

Tyr Leu Lys Thr Thr Arg Pro Phe Lys Thr Ser Ser Pro Ser Asn Leu
    130                 135                 140

Leu Gly Ala Lys Ile Leu Ser Val Val Ile Trp Ala Phe Met Phe Leu
145                 150                 155                 160

Ile Ser Leu Pro Asn Met Ile Leu Thr Asn Arg Arg Pro Lys Asp Lys
```

165 170 175

Asp Val Thr Lys Cys Ser Phe Leu Lys Ser Glu Phe Gly Leu Val Trp
180 185 190

His Glu Ile Val Asn Tyr Ile Cys Gln Val Ile Phe Trp Ile Asn Phe
195 200 205

Leu Ile Val Ile Val Cys Tyr Ser Leu Ile Thr Lys Glu Leu Tyr Arg
210 215 220

Ser Tyr Val Arg Thr Arg Gly Ser Ala Lys Val Pro Lys Lys Lys Val
225 230 235 240

Asn Val Lys Val Phe Ile Ile Ile Ala Val Phe Phe Ile Cys Phe Val
245 250 255

Pro Phe His Phe Ala Arg Ile Pro Tyr Thr Leu Ser Gln Thr Arg Ala
260 265 270

Val Phe Asp Cys Ser Ala Glu Asn Thr Leu Phe Tyr Val Lys Glu Ser
275 280 285

Thr Leu Trp Leu Thr Ser Leu Asn Ala Cys Leu Asp Pro Phe Ile Tyr
290 295 300

Phe Phe Leu Cys Lys Ser Phe Arg Asn Ser Leu Thr Ser Met Leu Arg
305 310 315 320

Cys Ser Asn Ser Thr Ser Thr Ser Gly Thr Asn Lys Lys Lys Gly Gln
325 330 335

Glu Gly Gly Glu Pro Ser Glu Glu Thr Pro Met
340 345

<210> 8
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: mouse oligonucleotide

<400> 8
cgttcctggg gttgataacc 20

<210> 9
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: mouse
oligonucleotide

<400> 9
cttgggaact ttggctgaac 20

<210> 10
<211> 356
<212> DNA
<213> Mus musculus

<400> 10
cgttcctggg gttgataacc attgaccgct acctgaagac caccaggcca tttaaaacgt 60
ccagccccag caatctcttg ggtgcaaaga ttctttctgt tgtcatctgg gccttcatgt 120
tcttaatttc actgcctaac atgattctca ccaacaggag gccaaaagat aaggacgtaa 180
caaaatgttc tttcttaaag tcagagtttg gtctagtttg gcacgaaata gtcaattaca 240
tctgccaagt cattttctgg attaattttt taattgtcat cgtttgttat agcctcatta 300
ccaaagaact ctatcggtct tatgtcagaa caaggggttc agccaaagtt cccaag 356